



	VICERRECTORADO DOCENTE	Código: GUIA-PRL-001
	CONSEJO ACADÉMICO	Aprobación: 2016/04/06
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		

		FORMATO DE INFORME DE PRÁCTICA DE LABORATORIO / TALLERES / CENTROS DE SIMULACIÓN – PARA ESTUDIANTES	
CARRERA: Computación		ASIGNATURA: Simulación	
NRO. PRÁCTICA:		TÍTULO PRÁCTICA: Examen de Interciclo	
OBJETIVO ALCANZADO:			
PREDICCIÓN DE TWEETS EN REDES SOCIALES			
<p>Primero se cargan las librerías necesarias para la ejecución de todo el programa</p> <pre>import pandas as pd import numpy as np from datetime import datetime from matplotlib import pyplot as plt from sklearn.model_selection import train_test_split as tts from sklearn.svm import SVC from sklearn.decomposition import TruncatedSVD from collections import Counter from nltk.corpus import stopwords from sklearn.metrics import mean_squared_error, mean_absolute_error</pre>			
<p>Para la primera parte se pide realizar el análisis de las redes sociales del actual presidente. En este caso se realizará el análisis de sus tweets en su cuenta oficial de Twitter https://twitter.com/LassoGuillermo. Para la extracción de la información se utilizó la herramienta Octaparse, que permite hacer web scraping de forma automática. Los datos recolectados se guardaron en el archivo Data.csv, y corresponden a:</p> <ul style="list-style-type: none"> • El usuario que realiza la publicación • La Fecha en la que se realiza la publicación • El texto del tweet publicado • El número de comentarios • El número de veces que la publicación fue compartida • El número de likes <p>Como primer paso se procede a aplicar técnicas Deprocesamiento de Lenguaje Natural (NPL) en la columna de los Tweets. En primer lugar se procede a eliminar información innecesaria como signos de puntuación (.,:) y saltos de línea (\n). Posteriormente se convierte todo el texto a minúsculas (Lowercase).</p> <ul style="list-style-type: none"> • Luego se realiza un análisis de la información para obtener nuevos datos de interés como: <p>El número de palabras que contiene la publicación La cantidad de etiquetas en la publicación El número de HashTags usados en la publicación Y finalmente se ordena la información obtenida en base al número de "Likes" de cada publicación (de menor a mayor).</p> <p>Otro dato interesante se obtiene a través del análisis de palabras utilizadas en cada Tweet. Por medio de la librería NLTK, que permite realizar NPL, se consigue una lista de las palabras más frecuentes utilizadas en todas las publicaciones. Para ello se omiten las Stopwords, que son palabras que carecen de información relevante, sin embargo tienden a repetirse más en la oración como los artículos (él, la, ellos). Una vez que se obtiene la lista de palabras más frecuentes se agrega una nueva columna al DataSet con la cantidad de veces que la publicación ha usado las palabras de esta lista.</p>			

	VICERRECTORADO DOCENTE	Código: GUIA-PRL-001
	CONSEJO ACADÉMICO	Aprobación: 2016/04/06
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		

	User	Date	Tweet	Comment	Share	Like	#Words	#HashTags	#Tags	Frequency	
0	Guillermo Lasso	10-Mar	garantizaremos la transparencia y mecanismos d...		16	17	70	28	0	0	1
1	AsíAmaneció	8-Mar	al aire periodistas ecuatorianas mujeres parti...		22	41	167	21	1	2	0
2	Red de Periodistas Libres #RPL	9-Mar	#buenasnoches a las 2000, miembros de la @rp...		14	42	178	27	6	3	1
3	Movimiento CREO Ecuador	9-Feb	el 135% del total de las actas representa a 1,...		43	55	230	34	1	0	1
4	Guillermo Lasso	11-Mar	encuentros como este enriquecen al país trabaj...		25	91	298	36	0	0	5
5	Josué Sanchez	4-Feb	¡#guayaquil votará por el equipo del cambiol e...		12	74	317	39	3	0	9
6	Guillermo Lasso	9-Mar	muchas gracias a los ganaderos que asistieron ...		25	64	326	25	1	0	4
7	Guillermo Lasso	23-Feb	muchas gracias a quienes se conectaron al en v...		41	39	351	19	0	0	2
8	Guillermo Lasso	11-Mar	hoy a las 17h00, #encontrémonosporlaruralidad ...		47	108	376	7	1	0	1
9	Guillermo Lasso	7-Feb	¡ahora! rueda de prensa sobre las elecciones d...		52	56	397	11	1	0	1
10	Cámara de Industrias	11-Mar	hoy nos acompaña el candidato a la presidencia...		26	106	408	21	0	1	2
11	Guillermo Lasso	8-Mar	la cooperación internacional debe de tener un ...		31	107	417	32	1	0	3
12	María Elena Machuca	3-May	presidente @lassoguillermo el camino está traz...		40	87	428	27	1	1	2
13	Guillermo Lasso	8-Mar	implementaremos mecanismos de protección, denu...		38	127	441	38	1	0	2
14	Guillermo Lasso	8-Mar	gracias a @wqradio_ec por la invitación a este...		50	128	442	33	1	1	2

Las columnas que se utilizarán como datos de entrada para nuestra regresión serán entonces:

- El número de palabras del tweet
- El número de HashTags usados
- La cantidad de etiquetas
- La cantidad de palabras frecuentes usadas
- La variable de salida que se intentará predecir será la cantidad de Likes que la publicación obtendrá.

Para el entrenamiento del sistema se divide entonces las variables de entrada y de salida en Train y Test para la etapa de entrenamiento y pruebas en un 80 y 20% respectivamente.

	#Words	#HashTags	#Tags	Frequency
293	19	1	0	4
169	28	1	0	3
133	6	0	0	0
203	1	2	0	2
265	26	0	1	0
...
68	36	3	0	4
266	40	2	1	3
111	40	0	0	5
9	11	1	0	1
127	39	0	0	1

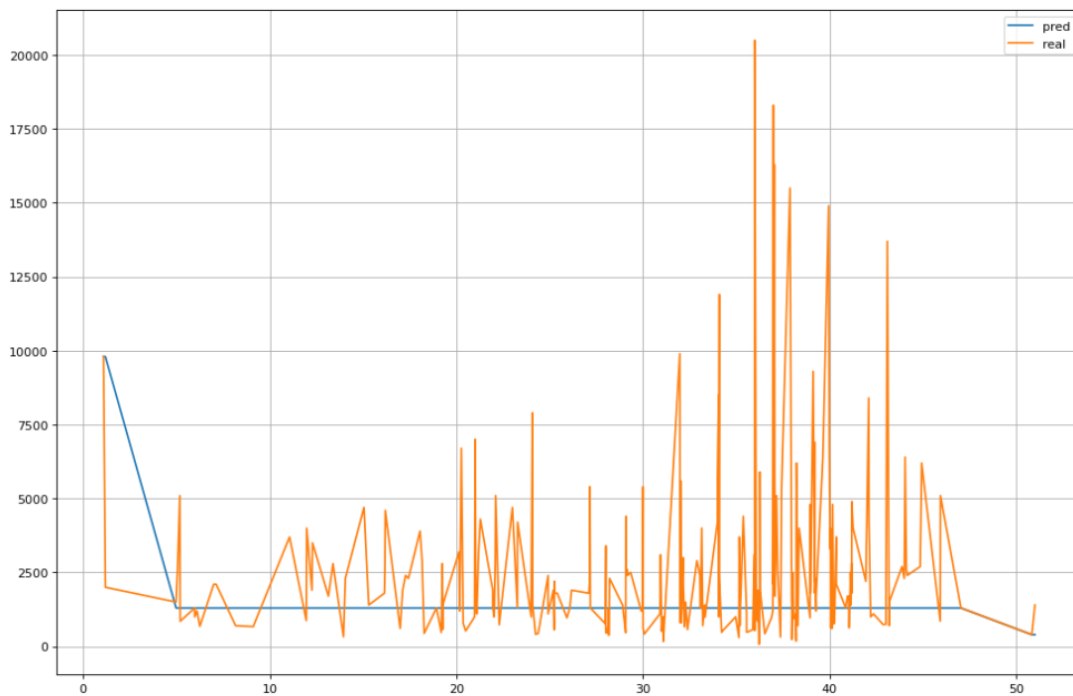
244 rows x 4 columns

Como se tienen cuatro parámetros de entrada, se realiza una reducción de dimensionalidad con el fin de tener una sola entrada y de esta forma realizar las gráficas respectivas en 2 dimensiones. Para esto se utiliza la técnica de SVD.

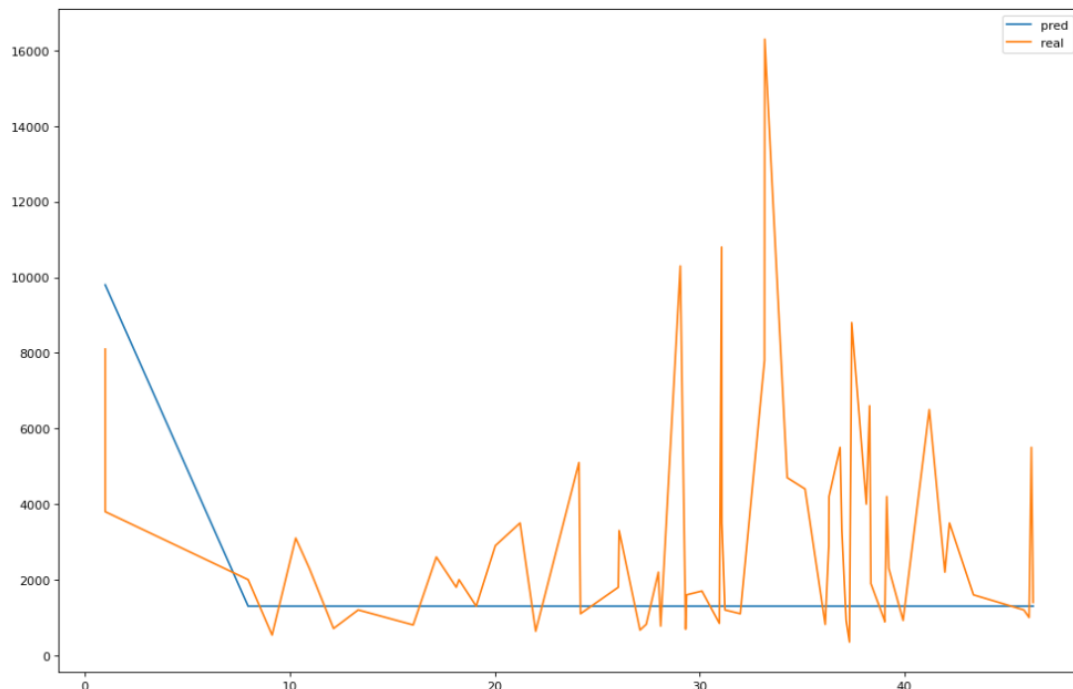
La regresión se realiza por medio de una máquina de vectores de soporte (SVM) que permite realizar un aprendizaje de máquina supervisado. En este caso se utilizó un kernel lineal.


```
clf=SVC(C=1,kernel="linear")
clf.fit(X_svd, TRlikes)
pred = clf.predict(X_svd)
```

En la gráfica se muestra la comparación de los datos reales vs. los predichos con los datos de entrenamiento.



Se procede a realizar la predicción de datos con los datos reservados previamente para las pruebas.



	VICERRECTORADO DOCENTE	Código: GUIA-PRL-001
	CONSEJO ACADÉMICO	Aprobación: 2016/04/06
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		

Como se puede ver en la gráfica, la predicción realizada por el algoritmo no es la mejor. Se puede decir que los parámetros de entrada escogidos no son los que definen cuál será la aceptación de la publicación. A continuación se presenta el Error Cuadrado Medio y el Error Absoluto Medio de las predicciones realizadas.

```
mse = mean_squared_error(TSlikes, pred)
mse
```

11921138.098360656

```
mae = mean_absolute_error(TSlikes, pred)
mae
```

2080.2950819672133

SIMULACIÓN DEL PROCESO DE VACUNACIÓN EN CUENCA

Primero se cargan las librerías necesarias para la ejecución de todo el programa

```
import simpy
import random
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
```

Se toma uno de los recintos en la parroquia El Batán de Cuenca. En esta zona hay 12738 personas que deberán votar en uno de los tres Recintos Electorales [1]. Suponiendo que todas las personas se reparten equitativamente en los 3 recintos, en cada recinto se atenderán a 4246 personas.

CANTÓN PARROQUIA	ÁREA*	ELECTORES	SEXO		TIPO DE VOTO		PERSONAS CON DISCAPACIDAD	JUNTAS	RECINTOS ELECTORALE
			HOMBRES	MUJERES	OBLIGATORIO	FACULTATIVO			
AZUAY		646.455	313.115	333.340	514.745	131.710	23.453	2.110	210
CAMILO PONCE ENRÍQUEZ		13.125	6.893	6.232	10.654	2.471	537	45	6
CAMILO PONCE ENRÍQUEZ	U	12.130	6.383	5.747	9.885	2.245	452	41	5
EL CARMEN DE PIJILÍ	R	995	510	485	769	226	85	4	1
CHORDELEG		11.627	5.613	6.014	9.275	2.352	466	38	6
CHORDELEG	U	8.232	4.098	4.134	6.660	1.572	297	26	2
LA UNIÓN	R	720	327	393	567	153	34	3	1
LUIS GALARZA O. (DELEGSOL)	R	864	356	508	656	208	46	3	1
PRINCIPAL	R	1.200	579	621	913	287	67	4	1
SAN MARTÍN DE PUZHIO	R	611	253	358	479	132	22	2	1
CUENCA		453.067	218.417	234.650	365.085	87.982	15.126	1.435	104
BAÑOS	R	22.824	11.212	11.612	18.775	4.049	566	71	4
BELLAVISTA	U	16.119	7.664	8.455	13.517	2.602	446	50	3
CAÑARIBAMBA	U	9.461	4.548	4.913	7.791	1.670	355	29	2
CHAUCHA / ANGAS	R	1.260	664	596	879	381	94	10	4
CHECA JIDCAY	R	4.119	2.039	2.080	3.319	800	138	14	1
CHIQUINTAD	R	5.125	2.538	2.587	4.068	1.057	349	16	1
CUMBE	R	6.457	3.027	3.430	4.096	1.461	245	10	2
EL BATÁN	U	12.738	6.229	6.509	10.647	2.091	307	39	3
EL VECINO	U	43.512	21.305	22.207	35.687	7.825	1.341	136	6
HERMANO MIGUEL	U	7.988	3.743	4.245	6.654	1.334	244	26	2
HUAYNACAPAC	U	17.614	8.323	9.291	14.039	3.575	735	55	3
LLACAO	R	3.790	1.812	1.978	2.974	816	144	12	1
MACHÁNGARA	U	8.195	4.013	4.182	6.952	1.243	226	28	4
MOLLETURO	R	6.102	3.088	3.014	4.715	1.387	361	29	10
MONAY	U	12.381	5.908	6.473	10.476	1.905	373	38	2

Puesto a que sólo el 80% de la población será vacunada, de las 4246 personas que deberían asistir al recinto, se espera que realmente asistan 3397 personas.

	VICERRECTORADO DOCENTE	Código: GUIA-PRL-001
	CONSEJO ACADÉMICO	Aprobación: 2016/04/06
Formato: Guía de Práctica de Laboratorio / Talleres / Centros de Simulación		

Puesto a que el horario de atención para la vacunación es de 08h00 a 17h00 [2], se define el tiempo de ejecución de la simulación diario en 540 (9 horas en minutos)

```

TIEMPO = 540
PUNTOS_VACUNACION0 = 6
PUNTOS_VACUNACION = 6
PUNTOS_VACUNACION2 = int(PUNTOS_VACUNACION/2)
DIAS = 100

vacunados1 = []
no_vacunados1 = []

s1 = 0
n1 = 0

vacunados2 = []
no_vacunados2 = []

s2 = 0
n2 = 0

```

Se crean los métodos para simular los siguientes procesos:

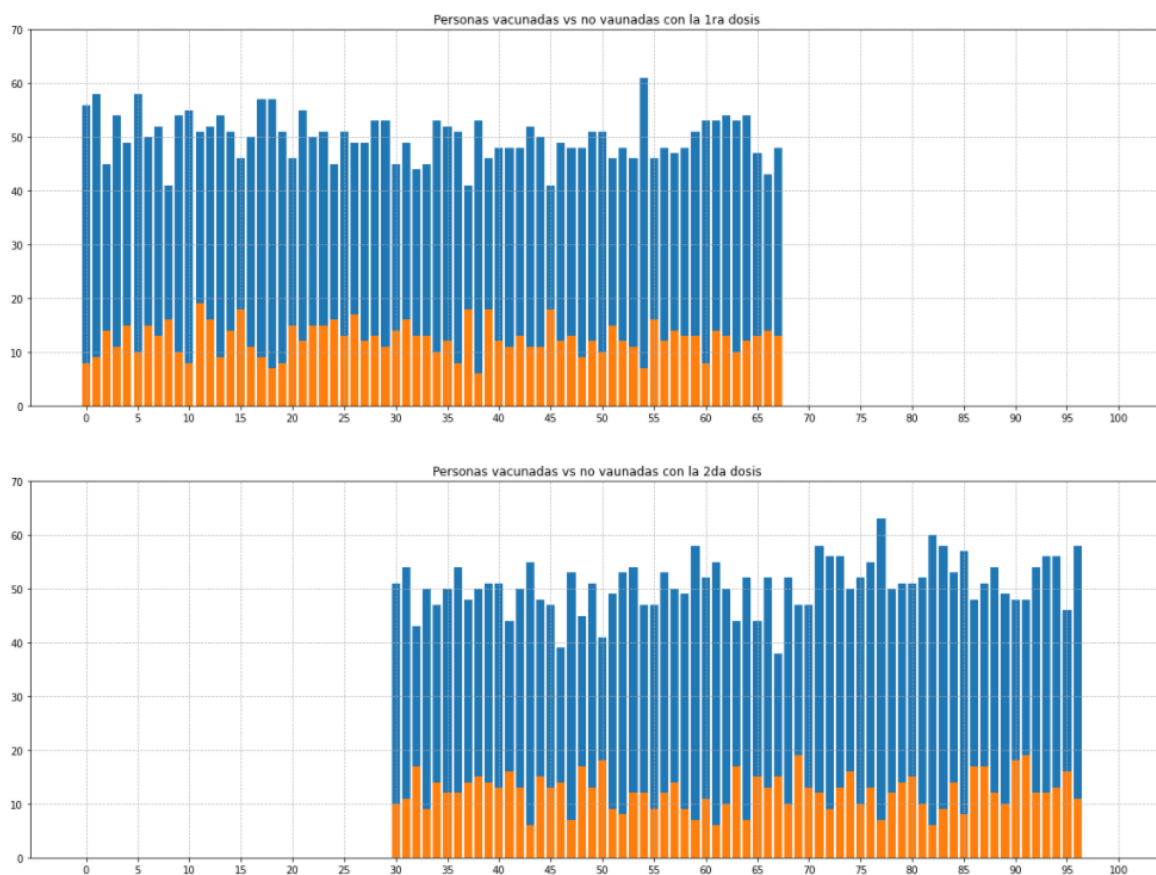
- Llegada del paciente con un tiempo de llegada entre 3 y 15 minutos
- Proceso de vacunación con una duración entre 5 y 10 minutos
- El tiempo de espera de 20 minutos
- Certificación con un tiempo de duración entre 2 y 3 minutos

Se ejecuta la simulación y se van mostrando en pantalla los procesos realizados por los pacientes que llegan a la vacunación.

----- DÍA 1 -----

El Paciente 1 llega al centro de vacunación a las 0.00. 1ra dosis
 El Paciente 2 llega al centro de vacunación a las 0.00. 1ra dosis
 El Paciente 3 llega al centro de vacunación a las 0.00. 1ra dosis
 El Paciente 4 llega al centro de vacunación a las 0.00. 1ra dosis
 El Paciente 5 llega al centro de vacunación a las 0.00. 1ra dosis
 ! El Paciente 5 no ha podido vacunarse la 1ra dosis y sale del recinto a las 0.00.
 El Paciente 6 llega al centro de vacunación a las 0.00. 1ra dosis
 > El Paciente 1 va a vacunarse a las 0.00. 1ra dosis
 > El Paciente 2 va a vacunarse a las 0.00. 1ra dosis
 > El Paciente 3 va a vacunarse a las 0.00. 1ra dosis
 > El Paciente 4 va a vacunarse a las 0.00. 1ra dosis
 > El Paciente 6 va a vacunarse a las 0.00. 1ra dosis
 > El Paciente 4 ha sido vacunado a las 6.00. 1ra dosis
 > El Paciente 6 ha sido vacunado a las 6.00. 1ra dosis
 _ El Paciente 4 comienza su tiempo de espera a las 6.00. 1ra dosis
 _ El Paciente 6 comienza su tiempo de espera a las 6.00. 1ra dosis
 > El Paciente 1 ha sido vacunado a las 7.00. 1ra dosis
 _ El Paciente 1 comienza su tiempo de espera a las 7.00. 1ra dosis
 > El Paciente 3 ha sido vacunado a las 8.00. 1ra dosis
 _ El Paciente 3 comienza su tiempo de espera a las 8.00. 1ra dosis
 El Paciente 7 llega al centro de vacunación a las 9.00. 1ra dosis
 > El Paciente 7 va a vacunarse a las 9.00. 1ra dosis
 > El Paciente 2 ha sido vacunado a las 10.00. 1ra dosis
 _ El Paciente 2 comienza su tiempo de espera a las 10.00. 1ra dosis
 El Paciente 8 llega al centro de vacunación a las 13.00. 1ra dosis
 > El Paciente 8 va a vacunarse a las 13.00. 1ra dosis
 > El Paciente 7 ha sido vacunado a las 18.00. 1ra dosis
 El Paciente 9 llega al centro de vacunación a las 18.00. 1ra dosis
 _ El Paciente 7 comienza su tiempo de espera a las 18.00. 1ra dosis
 > El Paciente 9 va a vacunarse a las 18.00. 1ra dosis
 > El Paciente 8 ha sido vacunado a las 21.00. 1ra dosis
 _ El Paciente 8 comienza su tiempo de espera a las 21.00. 1ra dosis
 > El Paciente 9 ha sido vacunado a las 24.00. 1ra dosis
 _ El Paciente 9 comienza su tiempo de espera a las 24.00. 1ra dosis
 _ El Paciente 4 termina su tiempo de espera a las 26.00. 1ra dosis
 @ El Paciente 4 va a recibir su certificado a las 26.00. 1ra dosis
 _ El Paciente 6 termina su tiempo de espera a las 26.00. 1ra dosis
 @ El Paciente 6 va a recibir su certificado a las 26.00. 1ra dosis
 El Paciente 10 llega al centro de vacunación a las 26.00. 1ra dosis
 > El Paciente 10 va a vacunarse a las 26.00. 1ra dosis
 _ El Paciente 1 termina su tiempo de espera a las 27.00. 1ra dosis
 @ El Paciente 1 va a recibir su certificado a las 27.00. 1ra dosis
 _ El Paciente 3 termina su tiempo de espera a las 28.00. 1ra dosis
 @ El Paciente 3 va a recibir su certificado a las 28.00. 1ra dosis
 @ El Paciente 6 ha recibido su certificado a las 28.00. 1ra dosis y sale del recinto
 @ El Paciente 4 ha recibido su certificado a las 29.00. 1ra dosis y sale del recinto
 @ El Paciente 1 ha recibido su certificado a las 29.00. 1ra dosis y sale del recinto
 _ El Paciente 2 termina su tiempo de espera a las 30.00. 1ra dosis
 @ El Paciente 2 va a recibir su certificado a las 30.00. 1ra dosis
 El Paciente 11 llega al centro de vacunación a las 30.00. 1ra dosis
 > El Paciente 3 ha recibido su certificado a las 30.00. 1ra dosis y sale del recinto
 > El Paciente 11 va a vacunarse a las 30.00. 1ra dosis
 @ El Paciente 2 ha recibido su certificado a las 32.00. 1ra dosis y sale del recinto
 > El Paciente 10 ha sido vacunado a las 33.00. 1ra dosis
 _ El Paciente 10 comienza su tiempo de espera a las 33.00. 1ra dosis
 > El Paciente 11 ha sido vacunado a las 36.00. 1ra dosis
 _ El Paciente 11 comienza su tiempo de espera a las 36.00. 1ra dosis
 _ El Paciente 7 termina su tiempo de espera a las 38.00. 1ra dosis
 @ El Paciente 7 va a recibir su certificado a las 38.00. 1ra dosis
 El Paciente 12 llega al centro de vacunación a las 40.00. 1ra dosis
 @ El Paciente 7 ha recibido su certificado a las 40.00. 1ra dosis y sale del recinto
 > El Paciente 12 va a vacunarse a las 40.00. 1ra dosis
 _ El Paciente 8 termina su tiempo de espera a las 41.00. 1ra dosis
 @ El Paciente 8 va a recibir su certificado a las 41.00. 1ra dosis
 El Paciente 13 llega al centro de vacunación a las 43.00. 1ra dosis
 ! El Paciente 13 no ha podido vacunarse la 1ra dosis y sale del recinto a las 43.00.
 _ El Paciente 9 termina su tiempo de espera a las 44.00. 1ra dosis
 @ El Paciente 9 va a recibir su certificado a las 44.00. 1ra dosis
 @ El Paciente 8 ha recibido su certificado a las 44.00. 1ra dosis y sale del recinto
 @ El Paciente 9 ha recibido su certificado a las 46.00. 1ra dosis y sale del recinto

Se muestran las gráficas de las personas vacunadas vs las no vacunadas para la primera y segunda dosis.



Nombre del estudiante: Doménica Merchán García