



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE INVESTIGACIONES EN MATEMÁTICAS
APLICADAS Y SISTEMAS

Visualización de la Información

Alejandro Hernández Rodríguez

Data Sexing

Índice general

0.1. Introducción	3
0.2. Fuente de datos	3
0.3. Gráficos	5

Índice de figuras

1.	Diagrama de pastel.	5
2.	Diagrama de dispersión hombres.	6
3.	Diagrama de pastel.	7
4.	Diagrama de pastel.	8

Índice de cuadros

1.	Datos antes de transformación.	4
2.	Datos después de la transformación.	4
3.	Columna Hermanos Edades.	4
4.	Columnas generadas de tipo de hermano.	5

0.1. Introducción

El sexo ha sido una de las principales interrogativas del humano durante su existencia. Desde tiempo antiguos hasta la modernidad filósofos, médicos y psicólogos han buscado entender que define la conducta sexual del humano. A busca de una respuesta de esta pregunta he decidido utilizar mis conocimientos de programación y ciencia de datos para poder obtener sino una respuesta al menos una idea de que factores pueden estar relacionados con la conducta sexual de algún individuo. Para esto decidí encuestar a mis colegas estudiantes mi universidad con un conjunto de preguntas básicas. Por ejemplo:

- Edad
- Escolaridad de sus padres
- Ubicación geográfica
- etc.

Esto con el objetivo de contestar una pregunta ¿Es posible predecir la frecuencia mensual de encuentros sexuales de un estudiante a partir de datos básicos utilizando algún algoritmo de Machine Learning? Entonces el plan fue el siguiente: Crear la encuesta, recorrer CU, limpiar datos, transformar datos, realizar analisis (visualización de los datos) y por último implementación del algoritmo de predicción.

0.2. Fuente de datos

Los datos que he obtenido fue a través de encuestas en línea utilizando la plataforma de google forms. Personalmente estuve visitando las distintas facultades de la universidad pidiendo a los alumnos que la contestaran mientras supervisaba que la contestaran de la manera correcta, más vale prevenir que limpiar. Lo hice de esa forma por qué considere que si la subía a redes sociales la cantidad de datos basura sería abismal. Lo útil de Google Forms es que una vez realizada las encuestas la plataforma genera automáticamente un google sheets donde representa cada pregunta cómo una columna y cada encuesta cómo una fila. Esto me permitió manipularlo muy facilmente utilizando el modulo de python pandas.

Facultad	Sexo	Edad	Semestre(número)	Orientación Sexual	Religión
Facultad de Ingeniería	Masculino	18	2	Heterosexual	Catolicismo
Facultad de Química	Femenino	18	2	Heterosexual	Catolicismo
Facultad de Química	Femenino	21	8	Heterosexual	Catolicismo

Cuadro 1: Datos antes de transformación.

Facultad	Sexo	Edad	Semestre(número)	Orientación Sexual	Religión
7	1	1	1	1	2
12	0	1	1	1	2
12	0	4	7	1	2

Cuadro 2: Datos después de la transformación.

La mayoría de las transformaciones fueron ordinales, es decir, codifique las variables categóricas con un mapeo a los números naturales. En la tabla 1 y 2 se observa una muestra de la codificación.

En lo que resultó en la siguiente tabla.

Otro ejemplo de transformación fue el de la columna que preguntaba la edad de los hermanos del estudiante. Este tenía que escribir sus edades separadas por comas. Entonces el dato estaba en formato de cadena.

Hermanos Edades
0
22
24,19,16
0

Cuadro 3: Columna Hermanos Edades.

Entonces para utilizar esta información lo que hice fue la creación de tres nuevas columnas que identificarán si el estudiante es el hermano menor, mayor o el de enmedio. Esto con el fin de observar si su rol familiar afecta su comportamiento sexual.

Hermano Menor	Hermano Enmedio	Hermano Mayor
0	0	1
1	0	0
0	1	0

Cuadro 4: Columnas generadas de tipo de hermano.

0.3. Gráficos

Primero tenemos que saber a quien estamos analizando y para saber eso, es útil utilizar gráficas descriptivas para poder realizar un análisis de los datos. Primero realice un diagrama de pastel para observar los porcentajes de estudiantes por cada facultad.

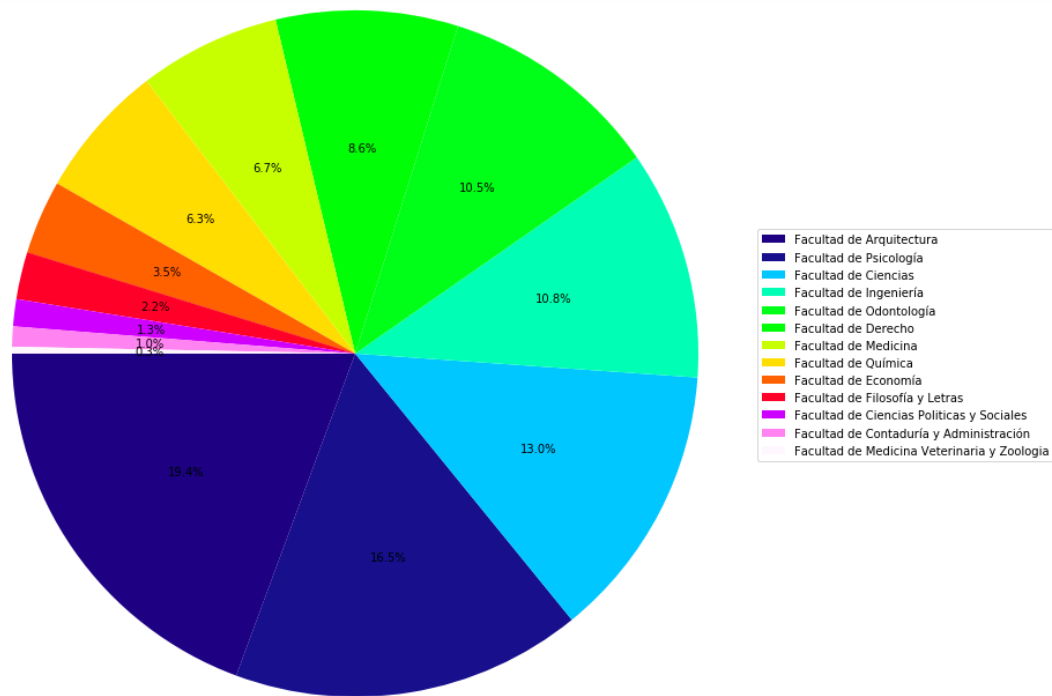


Figura 1: Diagrama de pastel.

Esta visualización me permitirá tomar decisiones futuras sobre que subpoblación considerar. Es fácil observar que no tomaré en cuenta la Facultad de Medicina Veterinaria y Zootecnia. Una vez segmentada la población es hora de comenzar el análisis para contestar la pregunta

de los 64 mil. El primero es visualizar si es que existe una relación entre el peso y la estatura con la actividad sexual. Para ello es fácil observarlo en un diagrama de dispersión.

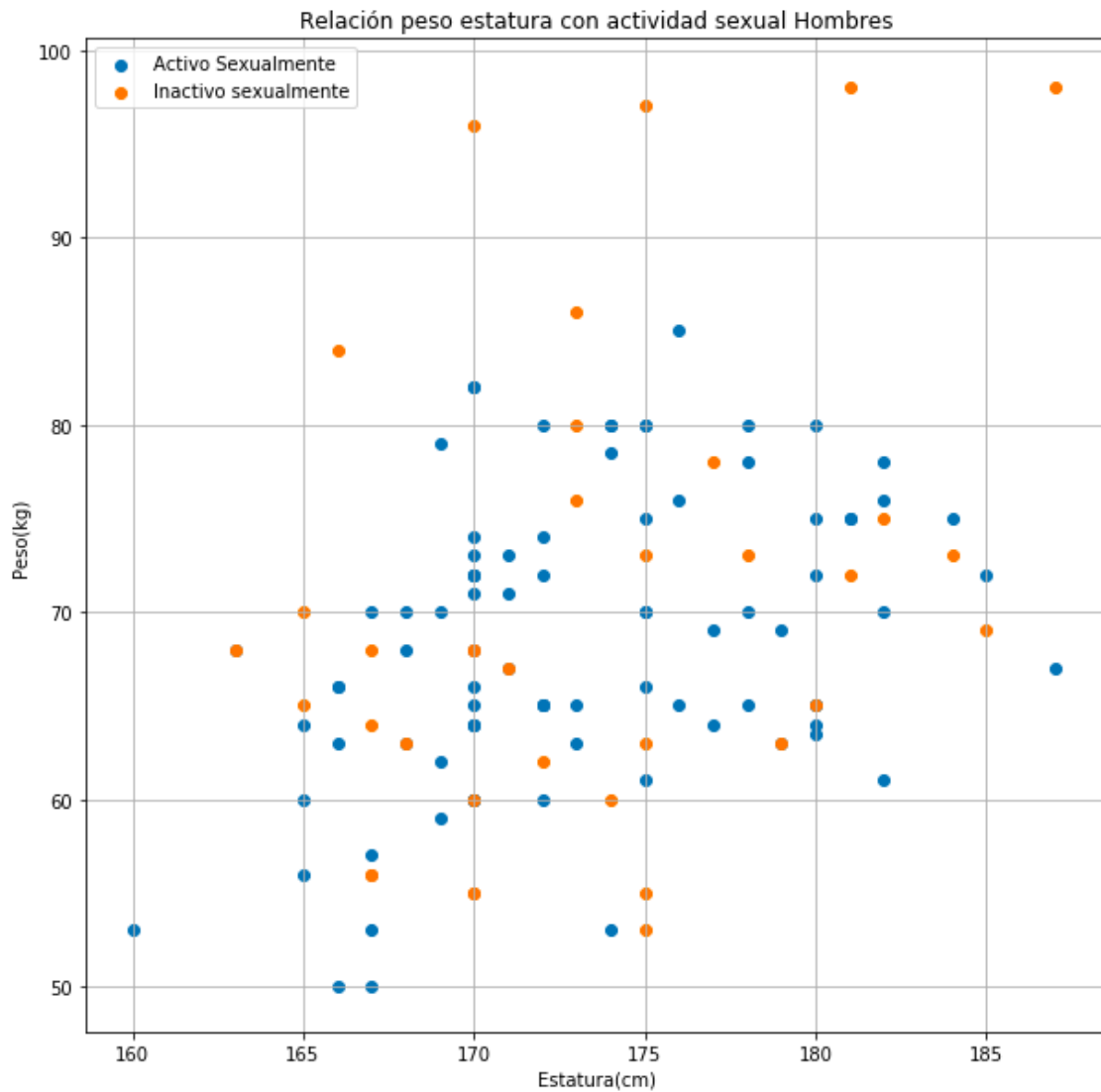


Figura 2: Diagrama de dispersión hombres.

Parece cómo si no tuvieran relación, pero se puede observar que las personas activas tienden a ser delgadas. Y esto es de alguna forma obvio, las personas con sobrepeso son menos atractivas. Siguiendo indagando en que factores pueden relacionarse con el comportamiento sexual. Decidí hacer otro par de gráficas sobre la frecuencia de relaciones sexuales mensuales. Primero dividiéndolo en hombres y mujeres como se muestra a continuación.

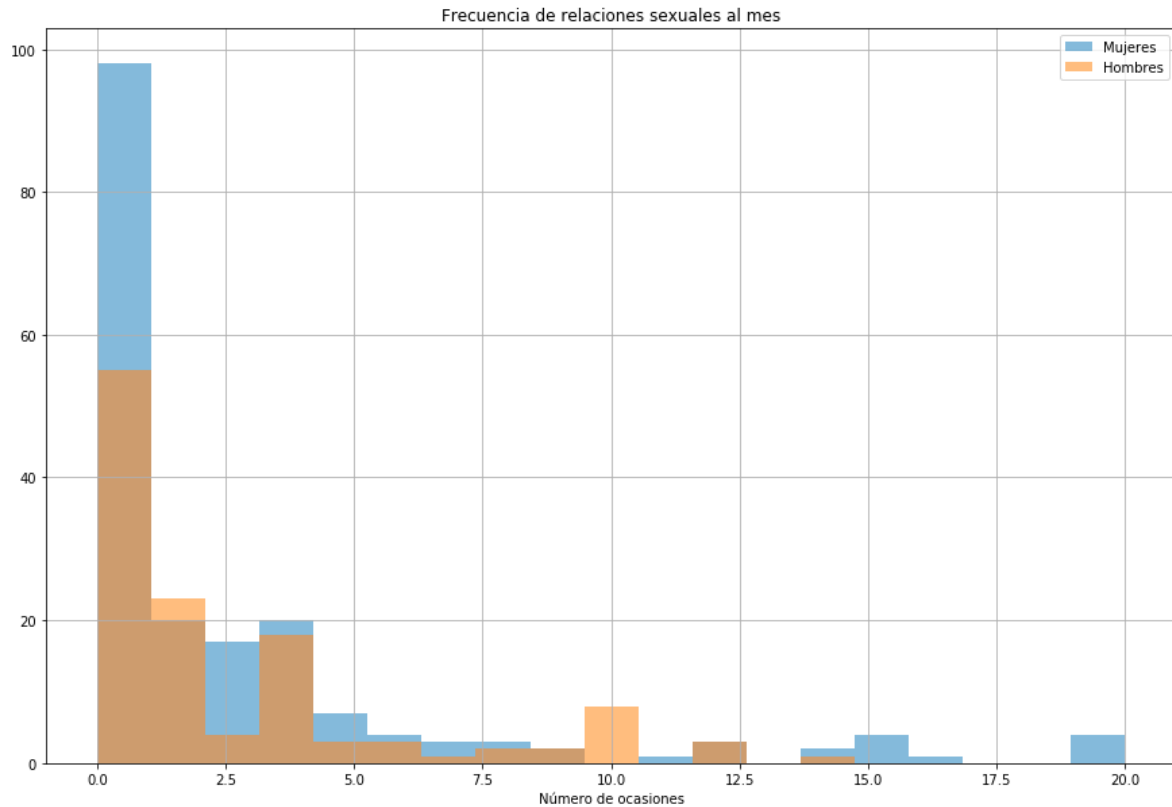


Figura 3: Diagrama de pastel.

Algo interesante de observar es que el conjunto femenino predomina en los extremos del histograma.

Después basandome en la gráfica de pastel anterior. Decidí utilizar las facultades cuyas cantidades de estudiantes encuestados eran similares. Para poder comparar con un histograma la cantidad de estudiantes con sus respectivas frecuencias de relaciones sexuales al mes.

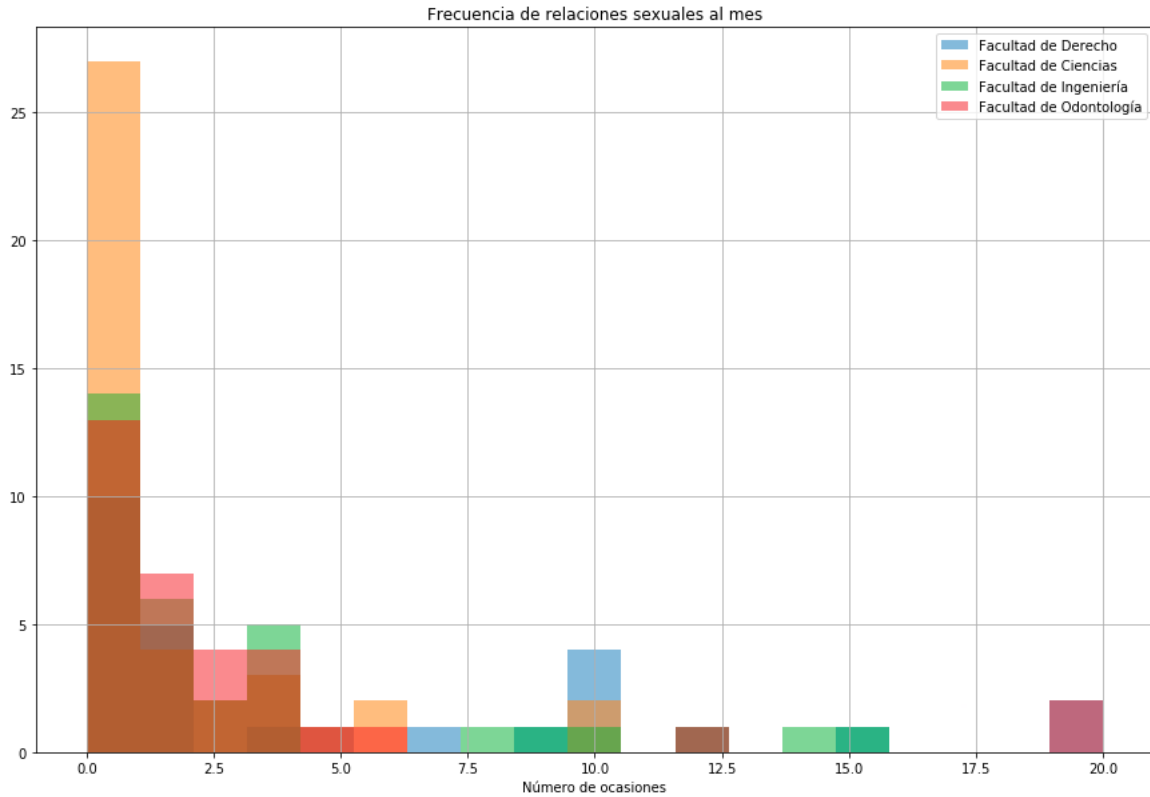


Figura 4: Diagrama de pastel.

Es impresionante como la Facultad de Ciencias rebasa por gran distancia las otra facultades en la primer intervalo del histograma que son de 0 a 2 encuentros sexuales al mes.

Las gráficas anteriormente descritas permitan tener una idea de la relación entre las distintas variables básicas de los estudiantes y su comportamiento sexual. Este es apenas un paso a un proyecto más desarrollado utilizando algoritmos estadísticos. Por el momento solo ganamos intuición y entendimiento acerca del comportamiento de este fenómeno a partir de la visualización de los datos.

Bibliografía

- [1] William B. Hansen, Electra D. Paskett, Linda J. Carter Health Education Research, Volume 14, Issue 4, August 1999, Pages 485–490, <https://doi.org/10.1093/her/14.4.485>