



Trabajo Final
ANÁLISIS DE DATOS CON PHYTON
VIDA UNIVERSITARIA

CREACIÓN DE UN SISTEMA MEDIANTE ALGORITMOS PHYTON

Castillo Sernaqué, Jairo Josué
Ceras Onofre, Yenny
Cornejo Taboada, Camila Fernanda
Fernández Palomino, Randy Joel
Nuñez Saravia, Fernando Martín

Universidad de Piura, 23 de enero de 2022

Esquema del Trabajo: Creación de un sistema para análisis de datos

1. Introducción.
2. Análisis del Sistema.
3. Ejemplos.
4. Conclusiones.

Análisis de Datos con Python

INTRODUCCIÓN

El presente trabajo utiliza una base de datos de 506 'communities areas' de Boston para hogares unifamiliares, el dataset contiene características asociadas a los inmuebles que se espera que tengan una relación con el precio de los inmuebles, tales como 'rooms' expresado en el promedio de habitaciones por inmueble, adicionalmente incluye variables externas que influyen en el precio como 'air pollution' la contaminación medida en óxido de nitrógeno por 100m., la distancia de la comunidad a los centros de empleo medida en la distancia ponderada al centro de empleo y el promedio del ratio estudiante-profesor en las escuelas locales 'stratio'. Los datos analizados corresponden a la base de individuos del año 2004.

Para el análisis de características de hogares unifamiliares se acota la muestra a 50 y que 9 características. Un aspecto importante es que mediante el programa vamos a poder aumentar las variables o las observaciones.

Se espera que el número promedio de habitaciones incremente el precio de los predios, y de acuerdo a los factores externos se espera que el precio promedio de los inmuebles en una determinada comunidad caiga a medida que ésta esté más contaminada, o que esté menos convenientemente situada a la disponibilidad de trabajos, o que posea escaso personal en sus escuelas, todo esto haría que los precios de los inmuebles no sean tan elevados de acuerdo a la oferta y demanda.

Además de analizar algunas de estas características el programa busca que dicha base de datos sea modificable en sus distintas variables, permite además poder realizar una actualización de los datos en la medida en que se recojan más información de otros predios, adicionalmente se puede realizar un análisis de datos filtrando los mismo por alguna determinada característica, genera gráficos estadísticos para una variable o para 2 variables, y permite adicionar nuevas características en base a las ya existentes.

Análisis de Datos con Python

1. Análisis del Sistema

El sistema en cuestión consiste en un programa donde se le permite al usuario realizar una serie de modificaciones en los datos permitiéndole actualizar la base de datos (es decir ingresar nuevas observaciones), generar nuevas características a partir de las ya existentes, acceder a un análisis gráfico y/o estadístico a fin de ver la relación de las características de los predios por 'areas communities' con el precio promedio de los mismos y establecer conclusiones.

En este sentido el sistema comienza preguntando si desea:

1. Ingresar nuevas observaciones y las guarda en el archivo excel.
2. Filtrar datos, ya sea por fila o por columna o también ordena toda la data dependiendo de la columna que escoja.
3. Graficar mediante histogramas cuando ingresamos una sola variable o un gráfico de dispersión cuando ingresamos dos variables.
4. Realiza medidas estadísticas de la columna que nosotros señalemos
5. Ingresamos nuevas columnas mediante tres opciones:
 - Ingresar el nombre de la nueva columna, da opción a que dicha columna se llene con números aleatorios donde ingresar un límite superior e inferior.
 - Ingresar el nombre de la nueva columna, da opción a que el usuario ingrese los valores para cada observación.
 - Ingresar el nombre de la nueva columna, da opción a que el usuario forme otra columna aplicando logaritmo neperiano a todos los elementos de dicha columna. Se hace esto porque muchas veces se utiliza para realizar regresiones en base a logaritmos neperianos de las variables.

Las opciones se repiten hasta que el usuario indique, que ya no desea realizar nada, por lo tanto presiona una tecla diferente a las indicadas.

Otra particularidad de este programa es que permite realizar filtración de datos para observar las particularidades de determinado grupo de individuos por categorías, las mismas que serán delimitadas por el usuario. En este sentido es el usuario quién debe definir si la filtración es por fila o por columna, el nombre de la variable a filtrar así como el orden ascendente o descendente en el que quiere que se muestren los datos.

Adicionalmente, resulta interesante poder generar resultados estadísticos mediante gráficos ya que permite al usuario entender de manera gráfica como se asocian las distintas variables y establecer conclusiones para la inversión en la industria inmobiliaria. Dado lo anterior, el sistema permite elegir al usuario qué gráfico desea construir mostrándole una lista de los distintos cuadros/gráficos estadísticos que se pueden construir, entre los que figura generar estadísticas de una determinada variable (*número de observaciones, media, desviación estándar, valores mínimos y*

Análisis de Datos con Python

máximos así como percentiles), mostrar gráficos estadísticos para una o dos variables, en el caso de una variable el usuario podrá replicar un histograma de la variable elegida y si son dos variables podrá obtener gráficos de dispersión de las mismas. Adicionalmente podrá elegir características particulares de efecto visual como el color con el que quiere que se visualicen los gráficos.

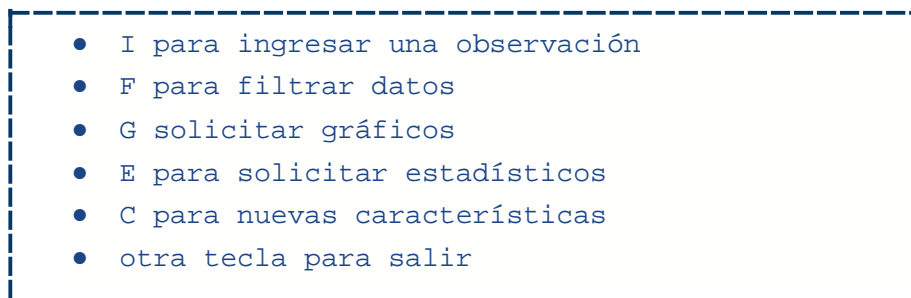
Para finalizar el sistema permite actualizar la base de datos con nuevas variables, el mismo está diseñado para preguntar de forma amigable qué característica desea agregar de manera que esta sería la nueva variable generada, además le da la opción de ingresar los datos de forma manual o aleatoria. En este último caso debe ingresarse los límites superiores o inferiores para generar los datos aleatoriamente, adicionalmente existe la posibilidad de querer logaritmar alguna de las variables existentes, esto último se creó para efectos de generar modelos económicos simples y sacar conclusiones a otro nivel.

2. Ejemplos

A continuación se muestran las opciones que el sistema le muestra al usuario a fin de que pueda elegir qué actividad desea realizar en el mismo.

Permitiéndole elegir agregar observaciones, filtración de datos, generar gráficos histograma o gráfico de dispersión, realizar estadísticos, agregar nuevas variables o salir del programa (ver gráfico 1)

Gráfico 1: Opciones de Usuario

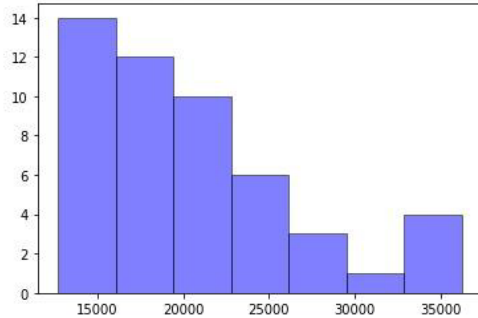
- 
- I para ingresar una observación
 - F para filtrar datos
 - G solicitar gráficos
 - E para solicitar estadísticos
 - C para nuevas características
 - otra tecla para salir

Para las opciones I, F el usuario debe elegir las variables y las características dispuestas por el programa (ver [video](#) para mayor explicación)

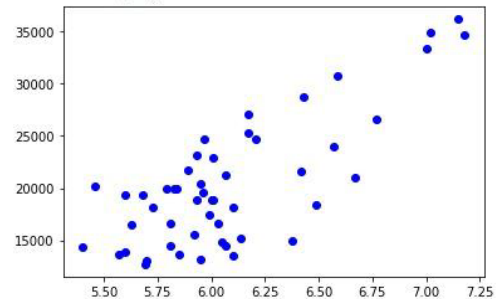
Por otro lado si el usuario busca generar gráficos estadísticos, este tiene dos opciones solicitar histogramas en el caso que quiera obtener información para una variable (ver gráfico 2) o si quiere hallar relación entre dos variables se puede generar un gráfico de dispersión en cuyo caso deberá elegir ambas variables a graficarse (ver gráfico 3).

Análisis de Datos con Python

**Gráfico 2: Histograma de la Variable
"Precio del Inmueble"**



**Gráfico 3: Scatterplot de "Habitaciones"
vs "Precio del Inmueble"**



En la siguiente dirección podrá observar la corrida del programa para las diferentes opciones:

<https://youtu.be/sRVDILJvA1g>

3. Conclusiones

De acuerdo a lo presentado líneas arriba podemos concluir que el sistema es un programa amigable que le permite al usuario realizar distintas alternativas: modificar la base de datos, realizar análisis mediante filtración de datos, generación de estadísticas así como establecer la relación entre las variables mediante gráficos.

Dado lo anterior el Sistema permite generar estadística descriptiva para la toma de decisiones en la industria inmobiliaria.

En la misma línea permite la actualización de datos y la incorporación de nuevos shocks que podrían afectar la industria. Por lo que es un sistema sostenible en el tiempo.

Permite realizar reportes estadísticos de utilidad para el inversor o para estudiar el efectos de los shocks en la industria inmobiliaria de manera que también se puede utilizar como material didáctico para explicar el comportamiento de dicho mercado en la academia.

Adicionalmente permite hacer inferencia estadística y sembrar nuevas inquietudes en los analistas para realizar modelos que permitan explicar el comportamiento de oferta y demanda en la industria inmobiliaria.

El sistema puede adecuarse a cualquier mercado inmobiliario incluso a cualquier otro tipo de mercado si se quisiera replicar.

Análisis de Datos con Python

Cabe mencionar que de acuerdo a la base de datos el sistema deja la posibilidad de poder realizar regresiones econométricas para ese mercado en particular. Queda pendiente el ejercicio para la siguiente etapa del curso.