

Entrega 1 - Proyecto IA

Luz Mariana Correa Pinto
Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano
Bogotá - Colombia
luzm.correap@utadeo.edu.co

Nathaly Fajardo Penagos
Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano
Bogotá - Colombia
nathaly.fajardop@utadeo.edu.co

Jeison Esteban Rojas Mora
Ciencias Naturales e Ingeniería
Universidad Jorge Tadeo Lozano
Bogotá - Colombia
jeison.rojasm@utadeo.edu.co

Resumen

Para esta primera entrega del proyecto de IA se utilizará la metodología de trabajo propuesta en el curso y las herramientas de modelamiento vistas para llevar a cabo la ejecución del proyecto (Precios del aguacate). Se realizará la extracción, pre-procesamiento, visualización y análisis de los datos obtenidos. Además, se encontrarán las principales características estadísticas de estos utilizando las herramientas vistas durante el curso.

Palabras clave:

Precios, aguacate, ventas, mercado, EE.UU, región, precio promedio, PLU 4046, PLU 4225, PLU 4770

1. Marco teórico

Problemática: Evaluar los datos históricos sobre los precios del aguacate y el volumen de ventas en varios mercados de EE.UU

Población de estudio: Personas nacidas entre 1981 y 1996

Lugar de estudio: Estados Unidos

Duración de la investigación: 3 años

2. Resultados pre-procesamiento y análisis de datos

En la **Figura.1a** se puede apreciar las librerías importadas (numpy, matplotlib, pandas, pylab, sklearn) con el fin de cargar los datos de la tabla “avocado.csv” usando la función `.read_csv` de la librería numpy. También se calculan los principales datos estadísticos de precio promedio (Average Price) de los aguacates. Podemos evidenciar:

- **Media:** 1.41
- **Mediana:** 1.37
- **Moda:** 1.15
- **Varianza:** 0.16

En la **Figura.1b** se muestran las estadísticas del total de aguacates vendidos (Total Volume):

- **Media:** 850644.01
- **Mediana:** 107376.76
- **Moda 1:** 2038.99
- **Moda 2:** 2858.31
- **Moda 3:** 3288.85
- **Moda 4:** 3529.44
- **Moda 5:** 3713.49
- **Moda 6:** 4103.97
- **Moda 7:** 7223.46
- **Moda 8:** 9465.99
- **Moda 9:** 13234.04
- **Moda 10:** 19634.24
- **Moda 11:** 46602.16
- **Moda 12:** 569349.05

- **Varianza:** 11926975521801.33

En la Figura.1c se muestran todos los datos estadísticos de cada uno de los campos que contiene el archivo “avocado.csv”. Entre estos datos estadísticos podemos evidenciar el total de datos, la media, desviación estándar, mínimo, máximo, primer cuartil, segundo cuartil y tercer cuartil de cada una de las columnas que componen el archivo

En la Figura.1d se observa el precio promedio del aguacate con una distribución más o menos normal y un precio promedio en 1.4

En la Figura.2a se observa el histograma con los datos de cada año

En la Figura.2b se muestra el histograma con los datos del total de aguacates vendidos

En la Figura.2c se muestra el histograma con los datos del total de aguacates con el PLU 4046

En la Figura.2d se muestra el histograma con los datos del total de aguacates con el PLU 40225

En la Figura.2e se muestra el histograma con los datos del total de aguacates con el PLU 4770

En la Figura.3a se muestra el histograma con los datos del total de aguacates hass

En la Figura.3b se muestra el histograma con los datos del total de aguates pequeños hass

En la Figura.3c se muestra el histograma con los datos del total de aguates grandes hass

```
colab.research.google.com/drive/124W3XoFGDRSAG3Z2ALAKT621dV1oFscrollTo=qHF-IHp7kUF
PrimeraEntregaProyectoIA.ipynb
Archivo Editor Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios
+ Código + Texto
versiones 4.4.0 - nuestro usar de algoritmos con CPU 4.4.0 versiones 4.7.0 - nuestro usar de algoritmos con CPU 4.7.0 versiones
#Se importan librerias
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import pylab as pl
from sklearn import preprocessing
from sklearn import datasets

#Se leen los datos
datos=pd.read_csv('avocado.csv')
#datos.columns

[4] # Se muestran las estadísticas del precio promedio de un aguacate

media = datos['AveragePrice'].mean()
print("la media es")
print(round(media,2))
mediana = datos['AveragePrice'].median()
print("la mediana es")
print(round(mediana,2))
moda = datos['AveragePrice'].mode()
print("la moda es")
print(round(moda,2))
var = datos['AveragePrice'].var()
print("la variancia es")
print(round(var,2))

la media es
1.41
la mediana es
1.27
la moda es
0
dtype: float64
la variancia es
0.36
```

(a) Librerías, carga de datos y estadísticas del precio promedio de un aguacate

```
colab.research.google.com/drive/124W3XoFGDRSAG3Z2ALAKT621dV1oFscrollTo=qHF-IHp7kUF
PrimeraEntregaProyectoIA.ipynb
Archivo Editor Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios
+ Código + Texto
# Se muestran las estadísticas del total de aguacates vendidos
media = datos['TotalVolume'].mean()
print("la media es")
print(round(media,2))
mediana = datos['TotalVolume'].median()
print("la mediana es")
print(round(mediana,2))
moda = datos['TotalVolume'].mode()
print("la moda es")
print(round(moda,2))
var = datos['TotalVolume'].var()
print("la variancia es")
print(round(var,2))

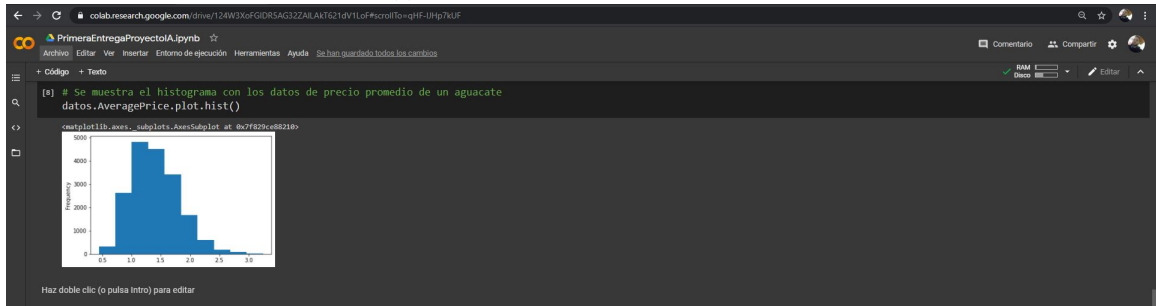
la media es
858644.81
la mediana es
187376.76
la moda es
0
0 2838.99
1 2838.99
2 1288.85
3 3529.44
4 3713.49
5 4181.97
6 7233.46
7 9465.99
8 13234.44
9 19634.34
10 46662.36
11 50249.45
dtype: float64
la variancia es
11926975521891.33
```

(b) Estadísticas del total de aguacates vendidos

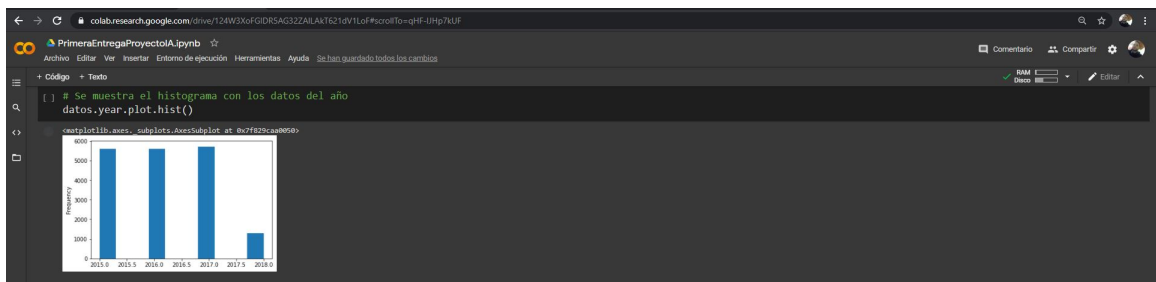
```
colab.research.google.com/drive/124W3XoFGDRSAG3Z2ALAKT621dV1oFscrollTo=qHF-IHp7kUF
PrimeraEntregaProyectoIA.ipynb
Archivo Editor Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios
+ Código + Texto
# Se muestran los datos estadísticos de cada columna
datos.describe()

Unnamed: 0  AveragePrice  TotalVolume  ag4946  ag4225  ag4770  TotalBags  SmallBags  LargeBags  XLargeBags  year
count  18249.000000  18249.000000  1.824900e+04  1.824900e+04  1.824900e+04  1.824900e+04  1.824900e+04  1.824900e+04  1.8249.000000
mean  24.232232  1.405978  0.506440e+05  2.930094e+05  2.283974e+04  2.396302e+05  1.821947e+05  6.433009e+04  3106.426507  2016.147899
std  15.481045  0.402677  3.453545e+06  1.264980e+06  1.204120e+06  9.862324e+05  7.461178e+05  2.439660e+05  17692.894652  0.939938
min  0.000000  0.440000  0.450200e+01  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00  2015.000000
25%  10.000000  0.440000  1.083850e+04  0.540700e+03  3.008780e+03  0.000000e+00  5.089640e+03  2.549420e+03  1.274710e+02  0.000000
50%  24.000000  1.370000  1.073760e+05  0.645300e+03  2.906103e+04  1.849990e+02  3.974383e+04  2.636202e+04  2.647710e+03  0.000000
75%  38.000000  1.660000  4.329623e+05  1.110203e+05  1.502069e+05  6.243420e+03  1.107834e+05  8.333767e+04  2.202050e+04  132.500000
max  52.000000  3.250000  6.250555e+07  2.274302e+07  2.047057e+07  2.546439e+06  1.937313e+07  1.338459e+07  5.719097e+06  551693.650000
```

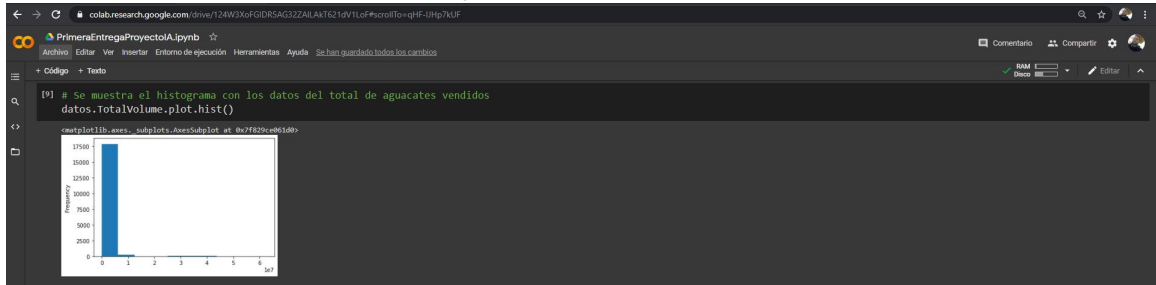
(c) Datos estadísticos de cada columna



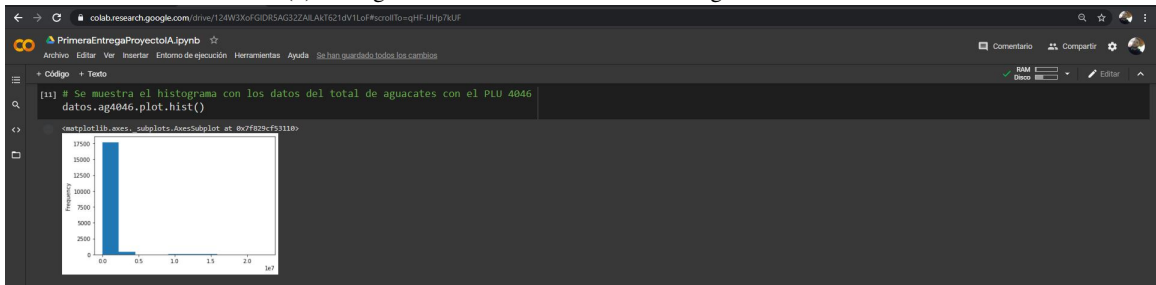
(d) Histograma con los datos de precio promedio de un aguacate



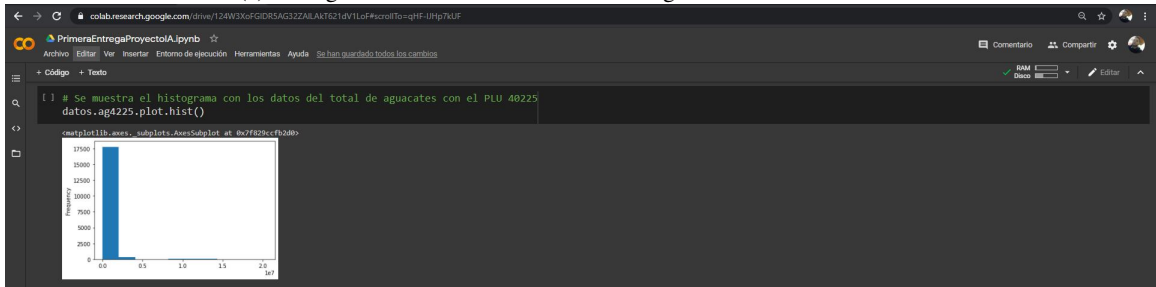
(a) histograma con los datos de cada año



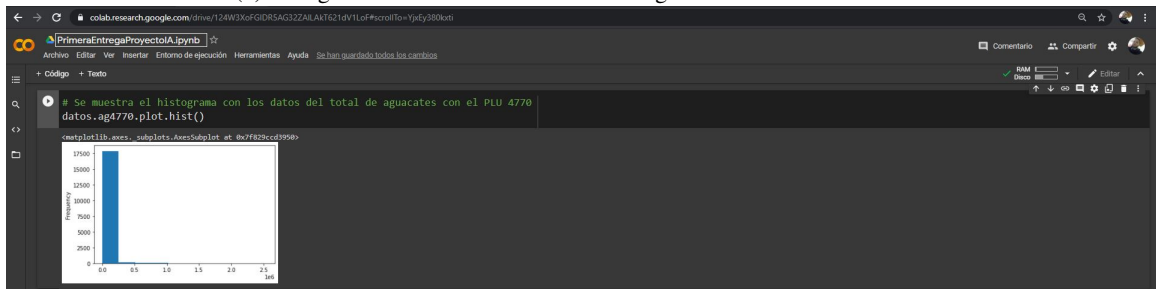
(b) Histograma con los datos del total de aguacates vendido



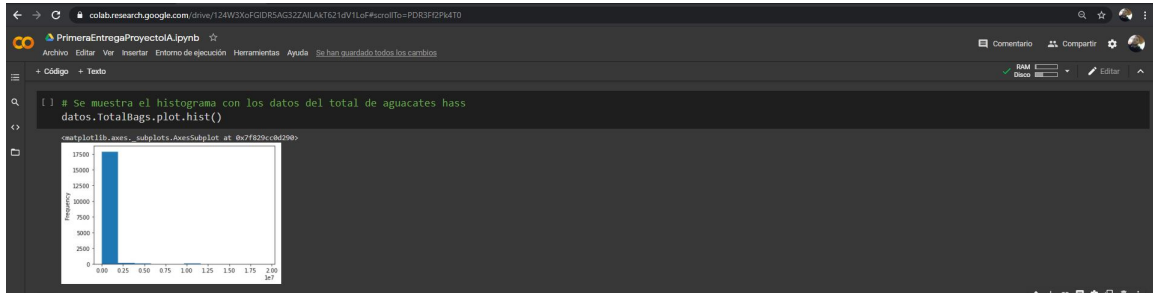
(c) Histograma con los datos del total de aguacates con el PLU 4046



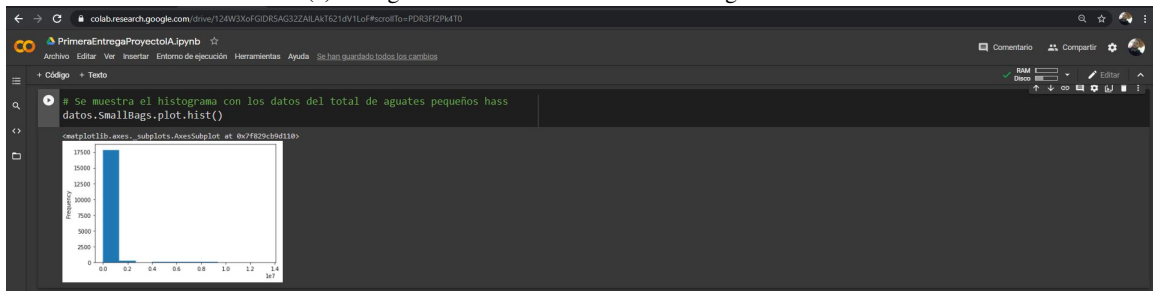
(d) histograma con los datos del total de aguacates con el PLU 40225



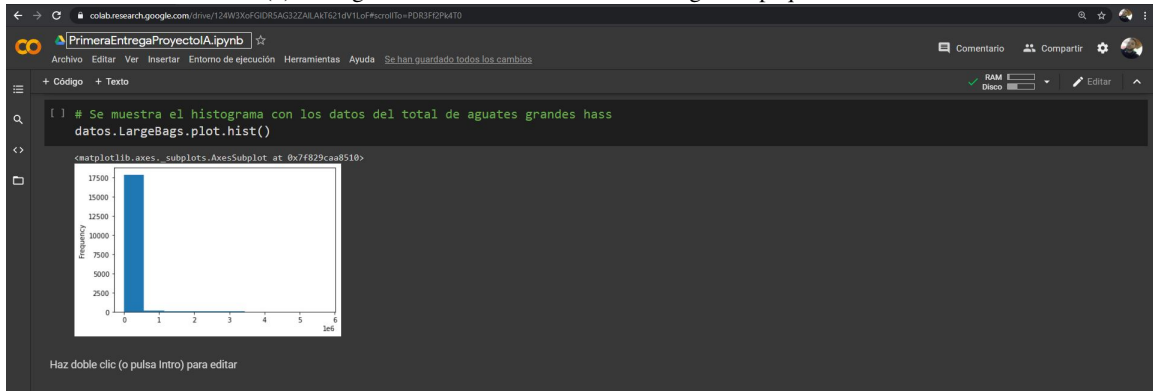
(e) Histograma con los datos del total de aguacates con el PLU 4770



(a) histograma con los datos del total de aguacates hass



(b) Histograma con los datos del total de aguates pequeños hass



(c) Histograma con los datos del total de aguates grandes hass