

1 PRACTICE 2

```
1 import streamlit as st
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 from sklearn.linear_model import LinearRegression
6 from sklearn.model_selection import train_test_split
7 from sklearn.metrics import mean_squared_error, r2_score
8
9 class DataHandler:
10     def __init__(self, file_path):
11         self.file_path = file_path
12         self.data = None
13
14     def load_data(self):
15         try:
16             self.data = pd.read_csv(self.file_path)
17             st.success("Datos cargados exitosamente.")
18         except Exception as e:
19             st.error(f"Error al cargar los datos: {str(e)}")
20
21     def preview_data(self):
22         if self.data is not None:
23             st.subheader("Vista previa de los datos")
24             st.dataframe(self.data.head(10))
25         else:
26             st.warning("No se han cargado datos.")
27
28     def calculate_statistics(self):
29         if self.data is not None:
30             st.subheader("Análisis estadístico")
31             numeric_columns = self.data.select_dtypes(include=['float64', 'int64']).columns
32             if len(numeric_columns) > 0:
33                 col1, col2 = st.columns(2)
34
35                 with col1:
36                     st.subheader("Media")
37                     for column in numeric_columns:
38                         mean_value = self.data[column].mean()
39                         st.write(f"{column}: {mean_value}")
40
41                     st.subheader("Mediana")
42                     for column in numeric_columns:
43                         median_value = self.data[column].median()
44                         st.write(f"{column}: {median_value}")
45
46                     st.subheader("Desviación estándar")
47                     for column in numeric_columns:
48                         std_value = self.data[column].std()
49                         st.write(f"{column}: {std_value}")
50
51                 with col2:
52                     st.subheader("Histogramas")
53                     fig, ax = plt.subplots(figsize=(5, 5))
54                     self.data.hist(ax=ax, bins=20)
55                     plt.tight_layout()
56                     st.pyplot(fig)
57
58             else:
59                 st.info("No hay columnas numéricas en el conjunto de datos.")
60         else:
61             st.warning("No se han cargado datos.")
62
63     def linear_regression(self):
64         if self.data is not None:
```

```

66         st.subheader("Regresi n Lineal ")
67         numeric_columns = self.data.select_dtypes(include=['float64', 'int64']).columns
68         if len(numeric_columns) >= 2:
69             target_column = st.selectbox("Selecciona la columna objetivo",
numeric_columns)
70             feature_columns = st.multiselect("Selecciona las columnas de
caracter sticas", numeric_columns.drop(target_column))
71
72             if len(feature_columns) > 0:
73                 X = self.data[feature_columns]
74                 y = self.data[target_column]
75                 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
76
77                 model = LinearRegression()
78                 model.fit(X_train, y_train)
79
80                 y_pred_train = model.predict(X_train)
81                 y_pred_test = model.predict(X_test)
82
83                 st.write("Coeficientes del modelo:", model.coef_)
84                 st.write("Intercepto del modelo:", model.intercept_)
85
86                 st.subheader("Evaluaci n del Modelo")
87
88                 col1, col2 = st.columns([2, 1])
89
90                 with col1:
91                     st.write("***Datos de entrenamiento**")
92                     st.write(f"Error Cuadr tico Medio (MSE): {mean_squared_error(
y_train, y_pred_train)}")
93                     st.write(f"Coeficiente de determinaci n (R ): {r2_score(y_train,
y_pred_train)}")
94
95                     st.write("***Datos de prueba**")
96                     st.write(f"Error Cuadr tico Medio (MSE): {mean_squared_error(y_test
, y_pred_test)}")
97                     st.write(f"Coeficiente de determinaci n (R ): {r2_score(y_test,
y_pred_test)}")
98
99                 with col2:
100                     st.subheader("Gr fico de Resultados")
101                     fig, ax = plt.subplots(figsize=(4, 4))
102                     ax.scatter(y_test, y_pred_test)
103                     ax.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=4)
104                     ax.set_xlabel('Valores reales')
105                     ax.set_ylabel('Valores predichos')
106                     st.pyplot(fig)
107             else:
108                 st.warning("Selecciona al menos una columna de caracter sticas.")
109         else:
110             st.warning("Se necesitan al menos dos columnas num ricas para realizar una
regresi n lineal.")
111         else:
112             st.warning("No se han cargado datos.")
113
114 def main():
115     st.set_page_config(page_title="An lisis de Datos CSV", layout="wide")
116
117     st.markdown(
118         """
119         <style>
120         .stApp {
121             background-color: #e6f7e6;
122         }
123         </style>
124         """,
125         unsafe_allow_html=True

```

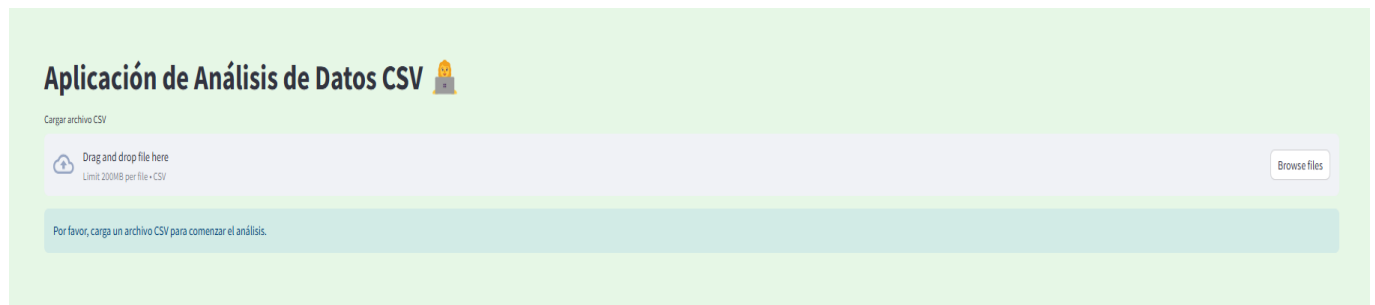
```

126 )
127
128 st.title("Aplicaci n de An lisis de Datos CSV")
129
130 uploaded_file = st.file_uploader("Cargar archivo CSV", type="csv")
131
132 if uploaded_file is not None:
133     data_handler = DataHandler(uploaded_file)
134     data_handler.load_data()
135
136     if data_handler.data is not None:
137         data_handler.preview_data()
138         data_handler.calculate_statistics()
139         data_handler.linear_regression()
140     else:
141         st.info("Por favor, carga un archivo CSV para comenzar el an lisis.")
142
143 if __name__ == '__main__':
144     main()

```

Listing 1: Código Python

1.1 Data Upload:



1.2 Data Visualization



1.3 Statistic analysis:



1.4 Statistic analysis:

Regresión Lineal

Selecciona la columna objetivo

Acidez

Selecciona las columnas de características

Alcohol x

Coefficientes del modelo:

value
0.0081

Intercepto del modelo: 2.9023653531685487

Evaluación del Modelo

Datos de entrenamiento

Error Cuadrático Medio (MSE): 0.3297960901684875

Coefficiente de determinación (R^2): 0.0006020585122777922

Datos de prueba

Error Cuadrático Medio (MSE): 0.3201047097705865

Coefficiente de determinación (R^2): -7.271861308177918e-05

2 Other CSV files

2.0.1 data set prime .csv

Aplicación de Análisis de Datos CSV

Cargar archivo CSV

 Drag and drop file here
Limit 200MB per file • CSV

 data-set-prime.csv 20.4KB

Datos cargados exitosamente.

Vista previa de los datos

	ID	Edad	Realizo_Ejercicio	Probabilidad
0	1	48	0	0.7
1	2	58	0	0.7
2	3	69	0	0.7
3	4	37	0	0.7
4	5	42	0	0.7
5	6	48	0	0.7
6	7	48	0	0.7
7	8	42	0	0.7
8	9	46	0	0.7
9	10	79	0	0.7
10	11	44	0	0.7

Análisis estadístico

Media

ID: 500.5

Edad: 48.401

Realizo_Ejercicio: 0.299

Probabilidad: 0.5804000000000001

Regresión Lineal

Selecciona la columna objetivo

ID

Selecciona las columnas de características

Edad 

Coefficientes del modelo:

value
-0.758

Intercepto del modelo: 535.4856447392185

Evaluación del Modelo

Datos de entrenamiento

Error Cuadrático Medio (MSE): 83521.27315436043

Coefficiente de determinación (R^2): 0.002280316698644036

Datos de prueba


Error Cuadrático Medio (MSE): 80528.74584019996


Coefficiente de determinación (R^2): 0.014709462245960925

2.0.2 datos logísticos.csv

Aplicación de Análisis de Datos CSV

Cargar archivo CSV

 Drag and drop file here
Limit 200MB per file • CSV

 datos-logisticos.csv 5.9KB

Datos cargados exitosamente.

Vista previa de los datos

	Muestra	Productos-Lote	Tiempo-Entrega	Defectuoso
0	1	19	19,000	0
1	2	35	20,000	0
2	3	26	43,000	0
3	4	27	57,000	0
4	5	19	76,000	0
5	6	27	58,000	0
6	7	27	84,000	0
7	8	32	150,000	1
8	9	25	33,000	0
9	10	35	65,000	0
10	11	26	80,000	0

Análisis estadístico

Media

Muestra: 200.5

Productos-Lote: 37.655

Tiempo-Entrega: 69742.5

Defectuoso: 0.3575

Regresión Lineal

Selecciona la columna objetivo

Muestra

Selecciona las columnas de características

Tiempo-Entrega ×

Coefficientes del modelo:

value
0.0006

Intercepto del modelo: 159.40440000773484

Evaluación del Modelo

Datos de entrenamiento

Error Cuadrático Medio (MSE): 12303.947107374875

Coefficiente de determinación (R^2): 0.03646146745698009

Datos de prueba

Error Cuadrático Medio (MSE): 15464.493946896395

Coefficiente de determinación (R^2): -0.0038085979374526957



