

Estimaciones Agrícolas

Proyecto Final



Cinthia Rodriguez

Comisión 61680
DATA SCIENCE- Machine
Learning para la Ciencia de
Datos

INDICE

• <i>Introducción</i>	<i>Pag. 3</i>
• <i>Objetivos</i>	<i>Pag. 3</i>
• <i>Análisis Comercial</i>	<i>Pag. 3</i>
• <i>Hipótesis</i>	<i>Pag. 3</i>
• <i>Pregunta de Interes</i>	<i>Pag. 4</i>
• <i>Objetivos-Aclaraciones</i>	<i>Pag. 5</i>
• <i>Métricas Utilizadas</i>	<i>Pag. 6</i>
• <i>Conclusiones</i>	<i>Pag.7</i>
• <i>Recomendaciones</i>	<i>Pag.8</i>



Introducción

Estimaciones Agrícolas es un documento de estadísticas agrícolas: cultivo/ cosecha/ localidades. Contiene información relacionada con la distribución zonal de semillas en el campo argentino. Este análisis nos permitirá identificar potenciales territorios para siembra/inversión.

OBJETIVO

- 1) Conocer las plantaciones más efectivas. Nos enfocaremos en el top 10 de todas las semillas, para potenciar las mismas.
- 2) Identificar futuras inversiones. En base a ese top 10, buscaremos herramientas agrarias que persuadan a los inversores extranjeros elegir el Territorio Argentino como primer opción de desembolso de dinero.
- 3) Detectar localidades en territorio argentino para la siembra. Argentina posee demográficamente localidades para potenciar la siembra y crecer a nivel mundial en cuestiones agrarias.

ANALISIS COMERCIAL

Destinado a:

- 1) Colaboradores administrativos
- 2) Jefes, Gerencia General, quien toma decisiones

HIPOTESIS

- 1) La Superficie Sembrada en este análisis no garantiza efectividad en la producción. Lo que se debe analizar también es el manejo de la tierra como un factor mucho mas relevante, ya que cada cierta cantidad de años se debe renovar. Cómo verificarlo: Comparar las superficies sembradas y cosechadas con la producción total mediante un gráfico de dispersión. Podría aplicarse un análisis de eficiencia para evaluar qué productores logran mejores rendimientos por hectárea.
- 2) El hecho de que una superficie brinde mejor producción no garantiza estabilidad económica, pues hay que considerar el factor climático, ya que puede destruir todo lo producido en un solo día y generar pérdidas irreparables.



Pregunta de Interés

- A- De donde se obtiene la información agraria?
- B- Cual es el objetivo con este análisis?
- C- Se puede analizar los últimos años para proyectar los próximos 2 venideros?
- D- Están incluidas todas las semillas que puedan ser plantadas en el Territorio Argentino?
- E- Quienes pueden invertir en Argentina?
- F- Se precisa conocimiento para invertir? Que protocolos seguir?
- G- Es confiable el análisis?
- H- Qué garantía hay con el ROI?



OBJETIVO

1) Conocer las plantaciones más efectivas. Nos enfocaremos en el top 10 de todas las semillas, para potenciar las mismas.

1) Identificar futuras inversiones. En base a ese top 10, buscaremos herramientas agrarias que persuadan a los inversores extranjeros elegir el Territorio Argentino como primer opción de desembolso de dinero.

	superficie cosechada	Inversion divisa USD
count	49830.000000	49830.000000
mean	11229.157977	858.746420
std	25053.162067	1795.638103
min	0.000000	79.910000
25%	228.000000	79.910000
50%	1600.000000	79.910000
75%	9500.000000	79.910000
max	349600.000000	5000.000000

1) Detectar localidades en territorio argentino para la siembra. Argentina posee demográficamente localidades para potenciar la siembra y crecer a nivel mundial en cuestiones agrarias.

	count
municipio_nombre	
Baradero	633
San Pedro	630
San Nicolás	628
Coronel Suarez	625
Balcarce	616
...	...
Tres de Febrero	1
Lanus	1
BOLIVAR	1
San Isidro	1
Avellaneda	1

123 rows x 1 columns



METRICAS UTILIZADAS

Medidas Descriptivas

```
[ ] # Medidas descriptivas para la columna 'superficie cosechada'
mean_superficie_cosechada = df['superficie cosechada'].mean()
median_superficie_cosechada = df['superficie cosechada'].median()
mode_superficie_cosechada = df['superficie cosechada'].mode()[0] # Primer valor de la moda

# Medidas de dispersión
std_superficie_cosechada = df['superficie cosechada'].std()#desviacion estandar

var_superficie_cosechada = df['superficie cosechada'].var()#varianza

print(f"Media: {mean_superficie_cosechada}")
print(f"Mediana: {median_superficie_cosechada}")
print(f"Moda: {mode_superficie_cosechada}")
print(f"Desviación Estándar: {std_superficie_cosechada}")
print(f"Varianza: {var_superficie_cosechada}")
```

Media: 11229.157977122215
Mediana: 1600.0
Moda: 0
Desviación Estándar: 25053.162067300276
Varianza: 627660929.5704135

Regresion Random Forest

```
[5] !pip install scikit-learn
from sklearn.ensemble import RandomForestRegressor
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
import numpy as np
```

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.14.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)

```
data = pd.read_csv('/content/drive/MyDrive/DS II/Estimaciones Agricolas TPDATASCI csv.csv')
data['roi_porcent'] = data['roi_porcent'].str.rstrip('%').astype(float) /100

X = data[['roi_porcent']]
y = data['roi_porcent']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Error cuadratico medio: {mse}")
print(f"R-cuadratico: {r2}")
```

Metricas de Clasificacion

```
] from sklearn.linear_model import LinearRegression
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score

data = pd.read_csv('/content/drive/MyDrive/DS II/Estimaciones Agricolas TPDATASCI csv.csv')
data['roi_porcent'] = data['roi_porcent'].str.rstrip('%').astype(float) /100

X = data[['roi_porcent']]
y = data['roi_porcent']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

model = LinearRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Error cuadratico medio: {mse}")
print(f"R-cuadratico: {r2}")
```

Error cuadratico medio: 2.1843599075327436e-33
R-cuadratico: 1.0

Clustering Jerarquico

amos a aplicar Clustering Jerárquico para identificar patrones en la producción de ajo en el país. Esto nos permitirá analizar la similitud entre productores (Agricultores o el Estado Argentino) en función de la distancia euclidiana entre ellos, especificar un número fijo de grupos.

```
] import pandas as pd
pd = pd.read_csv('/content/drive/MyDrive/DS II/Estimaciones Agricolas TPDATASCI csv.csv')

from sklearn.preprocessing import StandardScaler
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster.hierarchy import dendrogram, linkage
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.cluster.hierarchy as sch

# Seleccionamos columnas numéricas para el clustering
df_num = pd.select_dtypes(include=['int64', 'float64'])

# Escalamos los datos
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_num.dropna())

# Dendrograma
plt.figure(figsize=(12, 6))
dendrogram(linkage(df_scaled, method='ward'))
plt.title('Dendrograma - Clustering Jerárquico')
plt.xlabel('Observaciones')
plt.ylabel('Distancia Euclidiana')
plt.show()
```



CONCLUSIONES

Según los datos arrojados podemos observar lo siguiente:

Detectamos que son 4 semillas más efectivas para la plantación, las cuales podrán atraer mejores inversiones dado el amplio basto terreno para su cultivo.

🌱 **AJO:** La localidad donde mayor cosecha hay es en 25 de Mayo, Provincia de Buenos Aires. Con inversiones locales provenientes del Estado Argentino. Arrojando un rendimiento del 10% el primer año, con un retorno financiero de 3333 USD.

🌱 **SOJA:** La localidad donde mayor cosecha hay es en Adolfo Gonzales Chaves, Provincia de Buenos Aires. Con inversiones Extranjeras provenientes del Inversor Sr. Sean King. Arrojando un rendimiento del 1% el primer año, con un retorno financiero de 5467 USD.

🌱 **CEBADA:** La localidad donde mayor cosecha hay es en Alberti, Provincia de Buenos Aires. Con inversiones Extranjeras provenientes de la Inversora Sra. Marilia Mendoza. Arrojando un rendimiento del 7% el primer año, con un retorno financiero de 1400 USD.

🌱 **TRIGO:** La localidad donde mayor cosecha hay es en Almirante Brown, Provincia de Buenos Aires. Con inversiones Locales provenientes del Inversor Sr. Wo Jinn. Arrojando un rendimiento del 9% el primer año, con un retorno financiero de 4000 USD.

Datos de Cultivo:				
	cultivo_nombre	municipio_nombre	latitud	longitud
0	Ajo	25 de Mayo	-35.4432	-60.1721
1	Soja	Adolfo Gonzales Chaves	-38.0481	-60.0744
2	Cebada	Alberti	-35.0279	-60.2851
3	Trigo	Almirante Brown	-34.8116	-58.2619
Datos Financieros:				
	Origen_Empresa	Productor_nombre	Inversion_moneda	Inversion_divisa USD \
0	Local	Estado Argentino	ARS	79.91
1	Extranjera	Sean King	USD	5000.00
2	Extranjera	Marilia Mendoza	USD	5000.00
3	Local	Wo Jinn	ARS	79.91
	roi_porcent			
0	10%			
1	1%			
2	7%			
3	9%			
Datos de Producción:				
	superficie_sembrada	superficie_cosechada	produccion en toneladas \	
0	3	3	10	
1	15	15	82	
2	5	5	7	
3	2	2	8	
	rendimiento USD	year_fecha		
0	3333.0	January-20		
1	5467.0	January-21		
2	1400.0	January-22		
3	4000.0	January-23		



RECOMENDACIONES

Para responder a las preguntas de interés podemos ver que es una buena Opción invertir en Territorio Argentino, sin tanto protocolo legal.

Los inversores deberían tener cierto conocimiento en donde colocar su Dinero, para esto, el Estado Argentino continúa capacitando a productores Agrícolas e Ingenieros Agrónomos para este trabajo. En conjunto, logran Atraer mas propuestas para el campo.

La información es confiable, pues es de público conocimiento a través del Portal del gobierno nacional, la cual cualquier persona con/ sin conocimiento Puede acceder a la misma.

El ROI analizado previamente puede garantizar hasta cierto punto los futuros Ingresos. Para preveer los acontecimientos climáticos/ pérdidas, el Estado Argentino ofrece diferentes propuestas en la contratación de seguros agrícolas.

