



Tecnológico de Monterrey

Actividad 4 (Regresión lineal y logística)

Escuela: Instituto Tecnológico de Estudios Superiores de Monterrey

Materia: Desarrollo de proyectos de análisis de datos (Gpo 301)

Nivel Académico: Profesional

Ciudad: Puebla

Autor:

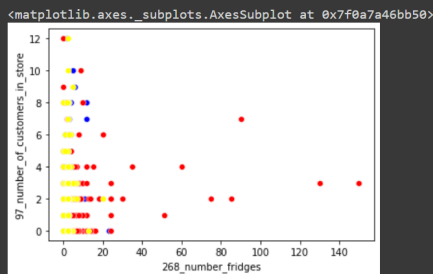
Juan Carlos Bejarano Moreno

A01732919

REGRESION LINEAL MULTIPLE:

```
[13] #Se grafican mapas de dispersión de las variables 3 variables: 'alcohol', 'speeding' y 'no_previous' al respecto a la variable "total"
from turtle import color

sns.scatterplot(x='268_number_fridges', y='97_number_of_customers_in_store', color="blue", data=datos)
sns.scatterplot(x='184_how_many_shelves_does_the_micro_retailer_have', y='97_number_of_customers_in_store', color="red", data=datos)
sns.scatterplot(x='183_number_own_fridges', y='97_number_of_customers_in_store', color="yellow", data=datos)
```

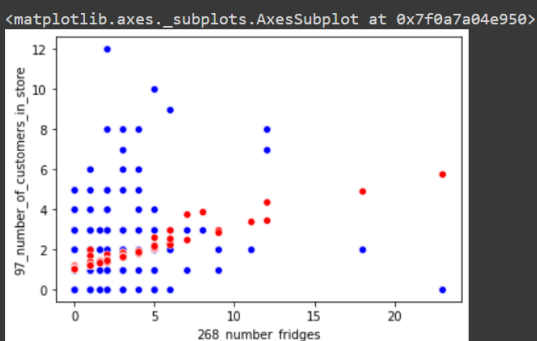


```
#Evaluamos la eficiencia del modelo obtenido por medio del coeficiente R2 Determinación
model.score(Vars_Indep, Var_Dep)
```

0.09041819227748749

```
#Visualizamos la gráfica comparativa entre el total real y el total predicho

sns.scatterplot(x='268_number_fridges', y='97_number_of_customers_in_store', color="blue", data=datos)
sns.scatterplot(x='268_number_fridges', y='Predicciones', color="red", data=datos)
#sns.lineplot(x='alcohol', y='Predicciones', color="red", data=df)
```



```
[24] #Corroboramos cual es el coeficiente de Determinación de nuestro modelo
coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
coef_Deter
```

0.09041819227748749

```
#Corroboramos cual es el coeficiente de Correlación de nuestro modelo
coef_Correl=np.sqrt(coef_Deter)
coef_Correl
```

0.30069617935299325

En base a las gráficas y valores finales de coeficientes de determinación y de correlación, podemos llegar a decir que dicho modelo puede ser confiable, hasta cierto punto. Aumentó considerablemente y lo mismo significa que dichas variables están relacionadas. La

inexactitud en este caso y en el siguiente se pueden generar por errores en los métodos de eliminación de valores nulos y outliers, afectado el desempeño general del modelo.

REGRESIÓN LOGÍSTICA:

```
#Calculo la precisión del modelo
from sklearn.metrics import precision_score

precision = precision_score(y_test, y_pred, average="weighted", pos_label="yes")
print("Precisión del modelo:")
print(precision)

Precisión del modelo:
0.8959648888391145
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1375: UserWarning: Note that pos_label (set to 'yes') is ignored when average != 'binary' (got 'weighted'). You may use labels=
UserWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use 'zero_divi
_warn_prf(average, modifier, msg_start, len(result))

#Calculo la exactitud del modelo
from sklearn.metrics import accuracy_score

exactitud = accuracy_score(y_test, y_pred)
print("Exactitud del modelo:")
print(exactitud)

Exactitud del modelo:
0.15151515151515152

[56] #Calculo la sensibilidad del modelo
from sklearn.metrics import recall_score

sensibilidad = recall_score(y_test, y_pred, average="weighted", pos_label="yes")
print("Sensibilidad del modelo:")
print(sensibilidad)

Sensibilidad del modelo:
0.15151515151515152
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1375: UserWarning: Note that pos_label (set to 'yes') is ignored when average != 'binary' (got 'weighted'). You may use labels=
UserWarning,
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Recall is ill-defined and being set to 0.0 in labels with no true samples. Use 'zero_division' pr
_warn_prf(average, modifier, msg_start, len(result))

# el puntaje F1 que es una combinación entre la precisión y la sensibilidad, para esto importamos f1_score.
#Calculo el Puntaje F1 del modelo
from sklearn.metrics import f1_score

puntajeF1 = f1_score(y_test, y_pred, average="weighted", pos_label="yes")
print("Puntaje F1 del modelo:")
print(puntajeF1)

Puntaje F1 del modelo:
0.8902389114548824
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1375: UserWarning: Note that pos_label (set to 'yes') is ignored when average != 'binary' (got 'weighted'). You may use labels=
UserWarning,
```

Una vez llegado al análisis de regresión logística, se revisaron las variables de:

97_number_of_customers_in_store

268_number_fridges

104_how_many_shelves_does_the_micro_retailer_have

Tomando como referencia de variable dependiente la de 104, se dejó como independiente las otras 2, a fin de que se pudiera revisar qué tanto impacto había en los mismos. Debido a los valores tan bajos presentes en dichas variables en su sensibilidad, podemos llegar a determinar que el valor y exactitud del modelo no es tan exacto y por lo mismo, no se puede llegar a una correlación entre las mismas variables y sus condiciones.