

```

# (Customer Churn) تحليل بيانات عملاء شركة اتصالات
# لتحليل سلوك العملاء

# استيراد جميع الأدوات التي سأتحتاجها 1.
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

# 2. رفع الملف يدوياً إلى Colab
from google.colab import files
print("...الرجاء اختيار ملف البيانات من جهازك")
uploaded = files.upload()

# 3. سيظهر اسم الملف الذي رفعته
# (في الغالب سيكون) استخدم اسم الملف الصحيح
file_name = list(uploaded.keys())[0]
df = pd.read_csv(file_name)

# 4. نظرة أولية سريعة
print("\n--- أول 5 صفوف في البيانات")
print(df.head())

print("\n--- معلومات عامة عن البيانات")
print(df.info())

print("\n--- إحصائيات سريعة")
print(df.describe())

# المرحلة الأولى: معالجة القيم المفقودة
# =====
print("\n" + "="*50)
print("المرحلة الأولى: معالجة القيم المفقودة")
print("="*50)

# 5. أكتشف عدد الخانات الفارغة في كل عمود
print("\nعدد القيم المفقودة قبل المعالجة:")
print(df.isnull().sum())

# 6. قد يكون نصياً، أحوله إلى رقمي 'TotalCharges' عمود
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')

# 7. بالمتوسط TotalCharges أعوض القيم المفقودة في
df['TotalCharges'].mean() = متوسط الفواتير
df['TotalCharges'].fillna(متوسط الفواتير, inplace=True)
print(f"بمتوسط القيمة وهو TotalCharges تم تعويض فراغات {متوسط الفواتير}")

print("\nعدد القيم المفقودة بعد المعالجة")
print(df.isnull().sum())

```

```

# المرحلة الثانية: معالجة القيم الشاذة
# =====
print("\n" + "="*50)
print("المرحلة الثانية: معالجة القيم الشاذة")
print("="*50)

الأعمدة_الرقمية = ['tenure', 'MonthlyCharges', 'TotalCharges']

def IQR_استبعاد_الشواذ_بطريقة(الجدول, الأعمدة):
    for عمود in الأعمدة:
        Q1 = الجدول[عمود].quantile(0.25)
        Q3 = الجدول[عمود].quantile(0.75)
        IQR = Q3 - Q1
        الحد_الأدنى = Q1 - 1.5 * IQR
        الحد_الأعلى = Q3 + 1.5 * IQR
        الجدول = الجدول[(الحد_الأدنى < [عمود][الجدول]) & ([عمود][الجدول] < الحد_الأعلى)]
        print(f"قيمة شاذة في عمود {عدد_الشواذ} تم العثور على")
        الجدول = الجدول[الجدول[عمود].where([عمود][الجدول] >= الحد_الأدنى)]
        الجدول.dropna(inplace=True)
    return الجدول

df = IQR_استبعاد_الشواذ_بطريقة(df, الأعمدة_الرقمية)
print(f"\nعدد الصفوف المتبقية بعد استبعاد الشواذ: {len(df)}")

# المرحلة الثالثة: التصورات البيانية
# =====
print("\n" + "="*50)
print("المرحلة الثالثة: التصورات البيانية")
print("="*50)

sns.set_style("whitegrid")
plt.rcParams['figure.figsize'] = (10, 6)

# الرسم الأول: توزيع مدة البقاء
plt.figure()
sns.histplot(df['tenure'], bins=20, kde=True, color='blue')
plt.title('الرسم 1: توزيع مدة بقاء العملاء (بالشهور)')
plt.xlabel('مدة البقاء')
plt.ylabel('عدد العملاء')
plt.show()

# الرسم الثاني: توزيع الفواتير الشهرية
plt.figure()
sns.histplot(df['MonthlyCharges'], bins=20, kde=True, color='green')
plt.title('الرسم 2: توزيع الفواتير الشهرية ($)')
plt.xlabel('الفواتير الشهرية ($)')
plt.ylabel('عدد العملاء')
plt.show()

# الرسم الثالث: نسبة العملاء الذين تركوا الخدمة
plt.figure()
df['Churn'].value_counts().plot(kind='pie', autopct='%1.1f%%', colors=['r', 'b'])
plt.title('الرسم 3: نسبة العملاء الذين تركوا الخدمة')
plt.ylabel('')
plt.show()

```

```

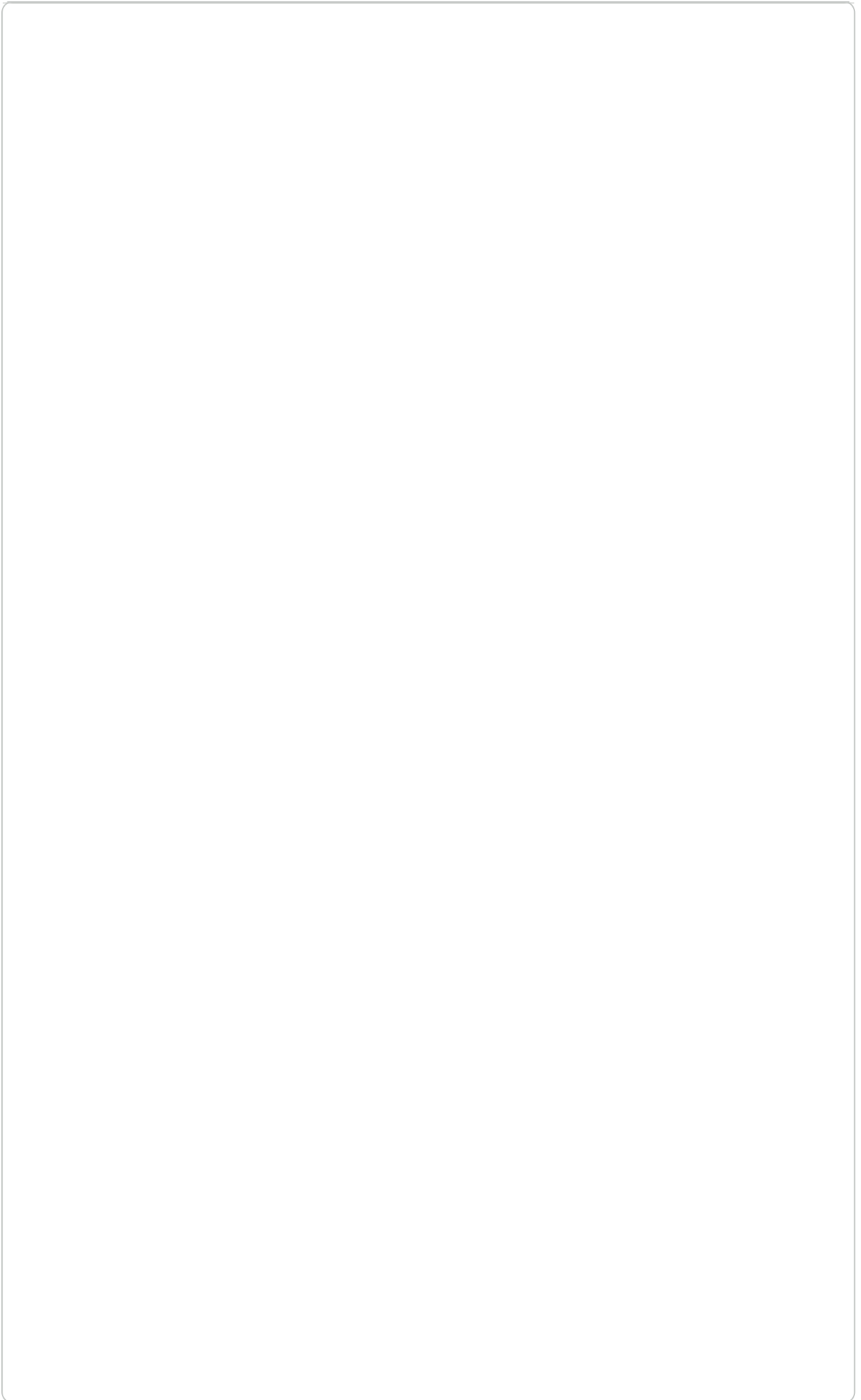
# الرسم الرابع: العلاقة بين مدة البقاء والترك
plt.figure()
sns.boxplot(x='Churn', y='tenure', data=df, palette='Set2')
plt.title('الرسم 4: مدة البقاء حسب حالة الترك')
plt.xlabel('هل ترك الخدمة؟')
plt.ylabel('مدة البقاء (شهر)')
plt.show()

# الرسم الخامس: خريطة الحرارة
df_encoded = df.copy()
for col in df_encoded.select_dtypes(include=['object']).columns:
    if col != 'customerID':
        df_encoded[col] = df_encoded[col].astype('category').cat.codes

plt.figure(figsize=(12, 10))
sns.heatmap(df_encoded.corr(), annot=False, cmap='coolwarm', linewidths=1)
plt.title('الرسم 5: خريطة ارتباط المتغيرات ببعضها')
plt.show()

# ملخص المشروع
# =====
print("\n" + "="*50)
print("ملخص مشروع:")
print("="*50)
print(f"1. عميل {len(df)}: حجم البيانات الأصلي")
print("2. TotalCharges بالمتوسط تم تحويل عمود")
print("3. تم استبعاد القيم الشاذة من الأعمدة الرقمية")
print("4. تم إنشاء 5 رسومات بيانية لفهم سلوك العملاء")
print("\nالملاحظات المستخلصة:")
print("- العملاء الذين بقوا فترة أطول أقل عرضة للترك")
print("- نوع العقد يؤثر بشكل كبير على قرار الترك")
print("- يمكن استخدام هذه الرؤى لتطوير استراتيجيات الاحتفاظ بالعملاء")

```



```
...الرجاء اختيار ملف البيانات من جهازك
اختيار الملفات WA_Fn-Use...mer-Churn.csv
WA_Fn-UseC_-Telco-Customer-Churn.csv(text/csv) - 977501 bytes, last modified:
10100% - 2026/7/ done
Saving WA_Fn-UseC_-Telco-Customer-Churn.csv to WA_Fn-UseC_-Telco-Customer-Cl

--- أول 5 صفوف في البيانات ---
  customerID  gender  SeniorCitizen  Partner  Dependents  tenure  PhoneService
0  7590-VHVEG  Female                0      Yes          No        1           No
1  5575-GNVDE   Male                0      No           No       34           Yes
2  3668-QPYBK   Male                0      No           No        2           Yes
3  7795-CFOCW   Male                0      No           No       45           No
4  9237-HQITU   Female              0      No           No        2           Yes

  MultipleLines  InternetService  OnlineSecurity  ...  DeviceProtection  \
0  No phone service          DSL                No  ...                No
1                No          DSL                Yes  ...                Yes
2                No          DSL                Yes  ...                No
3  No phone service          DSL                Yes  ...                Yes
4                No  Fiber optic                No  ...                No

  TechSupport  StreamingTV  StreamingMovies  Contract  PaperlessBilling
0          No          No          No  Month-to-month          Yes
1          No          No          No    One year          No
2          No          No          No  Month-to-month          Yes
3         Yes          No          No    One year          No
4          No          No          No  Month-to-month          Yes

  PaymentMethod  MonthlyCharges  TotalCharges  Churn
0  Electronic check          29.85          29.85    No
1    Mailed check          56.95         1889.5    No
2    Mailed check          53.85          108.15   Yes
3  Bank transfer (automatic)  42.30         1840.75    No
4  Electronic check          70.70          151.65   Yes

[5 rows x 21 columns]

--- معلومات عامة عن البيانات ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            7043 non-null  object
1   gender                 7043 non-null  object
2   SeniorCitizen          7043 non-null  int64
3   Partner                7043 non-null  object
4   Dependents             7043 non-null  object
5   tenure                 7043 non-null  int64
6   PhoneService           7043 non-null  object
7   MultipleLines           7043 non-null  object
8   InternetService         7043 non-null  object
9   OnlineSecurity          7043 non-null  object
10  OnlineBackup            7043 non-null  object
11  DeviceProtection        7043 non-null  object
12  TechSupport             7043 non-null  object
13  StreamingTV             7043 non-null  object
14  StreamingMovies         7043 non-null  object
15  Contract                7043 non-null  object
16  PaperlessBilling        7043 non-null  object
```

```

17 PaymentMethod      7043 non-null object
18 MonthlyCharges      7043 non-null float64
19 TotalCharges        7043 non-null object
20 Churn                7043 non-null object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
None

```

--- إحصائيات سريعة ---

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

=====

المرحلة الأولى: معالجة القيم المفقودة

=====

عدد القيم المفقودة قبل المعالجة:

```

customerID      0
gender           0
SeniorCitizen    0
Partner          0
Dependents       0
tenure           0
PhoneService     0
MultipleLines    0
InternetService  0
OnlineSecurity   0
OnlineBackup     0
DeviceProtection 0
TechSupport      0
StreamingTV      0
StreamingMovies  0
Contract         0
PaperlessBilling 0
PaymentMethod    0
MonthlyCharges   0
TotalCharges     0
Churn            0

```

dtype: int64

بمتوسط القيمة وهو: TotalCharges 2283.30 تم تعويض فراغات

عدد القيم المفقودة بعد المعالجة:

```

customerID      0
gender           0
SeniorCitizen    0
Partner          0
Dependents       0
tenure           0
PhoneService     0
MultipleLines    0
InternetService  0
OnlineSecurity   0
OnlineBackup     0

```

```

DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges        0
Churn               0
dtype: int64

```

=====

المرحلة الثانية: معالجة القيم الشاذة

=====

'tenure' تم العثور على 0 قيمة شاذة في عمود

'MonthlyCharges' تم العثور على 0 قيمة شاذة في عمود

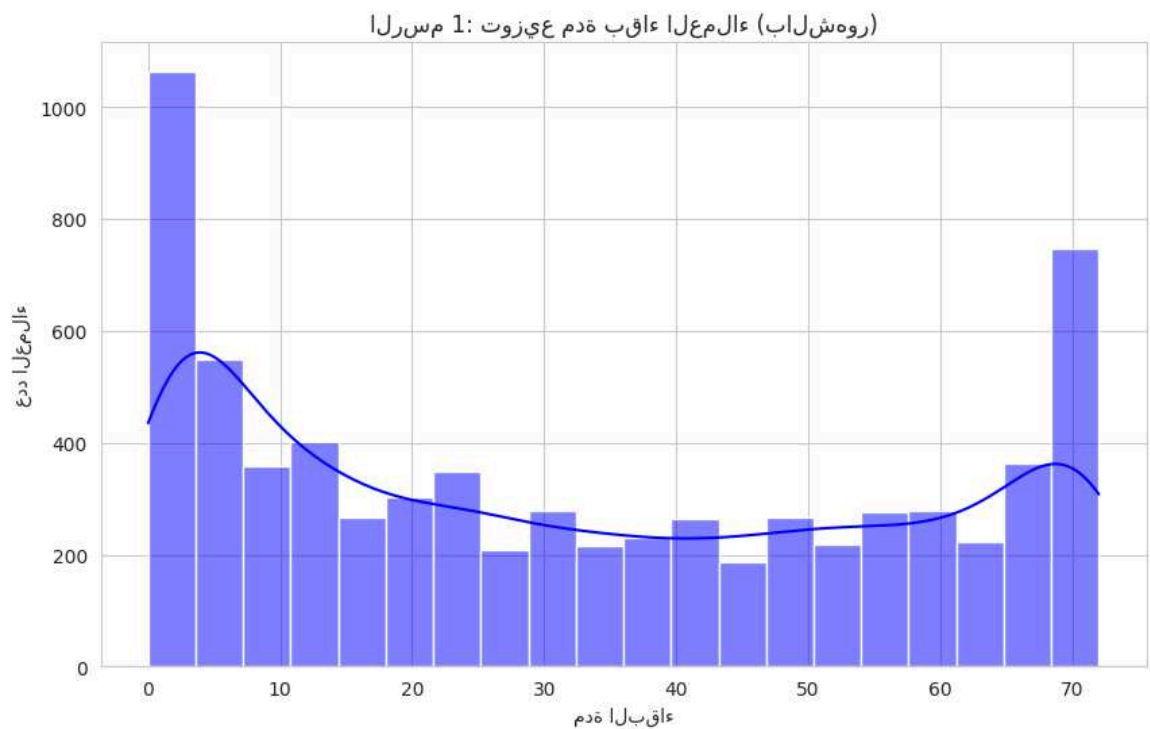
'TotalCharges' تم العثور على 0 قيمة شاذة في عمود

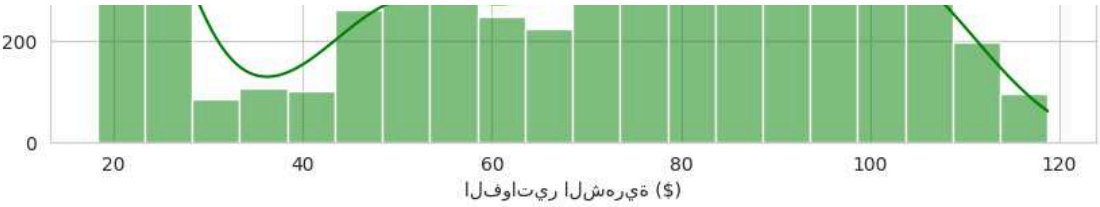
عدد الصفوف المتبقية بعد استبعاد الشواذ: 7043

=====

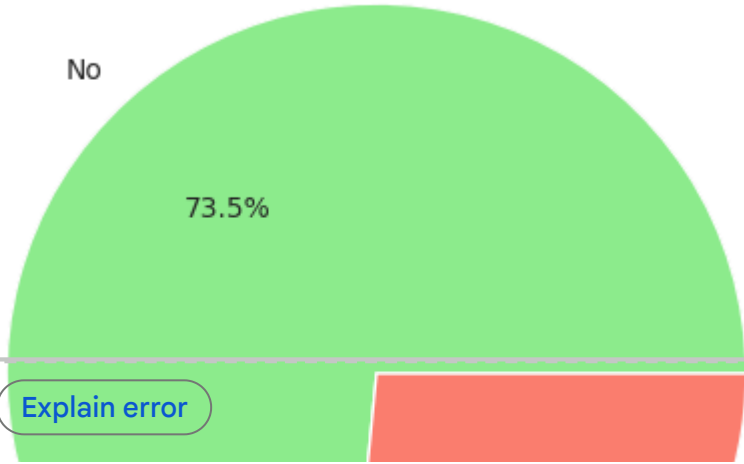
المرحلة الثالثة: التصورات البيانية

=====





مَدْخَالِ اَوَكْرَتِ نِيْذِلِ اءَالِ مَعَالِ اَقْبَسَن 3: مَسْرِلِ



Next steps: [Explain error](#)