Лабораторна робота №4. Варіант 15

Виконав: студент групи ІП-13 Ал Хадам Мурат Резгович

Візуалізація даних за допомогою matplotlib та Seaborn Мета роботи: Ознайомитись з основними діаграмами та графіками, що

використовуються при аналізі даних. Навчитись будувати їх за допомогою бібліотек

matplotlib та Seaborn.

Завдання до лабораторної роботи: 1. Побудувати стовпчикові діаграми, на яких відобразити: а) кількість клієнтів з підключеною голосовою поштою та без; б) мінімальну кількість хвилин в вечірній час для клієнтів з підключеною голосовою поштою і без; в) середню оплату

In [2]:

1000

0

міжнародних дзвінків для клієнтів з підключеним роумінгом та без і з врахуванням лояльності. 2. Побудувати гістограму хвилин в нічний час, загальну і в залежності від підключеного роумінгу. 3. Побудувати діаграму розмаху дзвінків в денний час (загальну і в залежності від підключеної голосової пошти), визначити чи присутні викиди. 4. За допомогою діаграм розсіювання зробити висновки щодо залежності між: а) міжнародними хвилинами і дзвінками; б) міжнародними хвилинами і оплатою. Порахувати коефіцієнт кореляції за допомогою відповідних функцій.

In [1]: import pandas as pd import matplotlib.pyplot as plt import seaborn as sns df = pd.read_csv('telecom.csv') df

Out[1]: number total total voice total total total total account area phone international

total

state mail vmail day day day eve eve night night number length code plan minutes plan messages minutes calls charge calls charge calls 382-0 KS 415 45.07 99 16.78 128 25 265.1 110 244.7 91 no yes 4657 371yes 1 ОН 107 415 26 161.6 123 27.47 103 16.62 254.4 103 no 7191 358-2 415 0 10.30 NJ 137 243.4 114 41.38 110 162.6 104 no no 1921 375-3 ОН 84 408 yes 0 299.4 71 50.90 88 5.26 196.9 89 no 9999 330-12.61 4 OK 415 0 166.7 28.34 122 186.9 121 75 113 yes no 6626 397-

1730 NM 161 408 0 107.5 121 18.28 46 21.79 247.2 no no 8011 341-1731 TX 70 415 0 232.8 95 39.58 111 25.79 255.6 no 8719 408-1732 ΗΙ 97 415 43 105 20.59 115 22.12 222.4 yes 121.1 no 1242

406-23.01 239.5 WA

131 104 100 7726 351-1734 92 415 157.7 101 26.81 ... 100 25.38 216.9 99 no 2773 1735 rows × 21 columns

1. Побудувати стовпчикові діаграми, на яких відобразити: а) кількість клієнтів з підключеною голосовою поштою та без;

plt.xlabel('Голосова пошта підключена') plt.ylabel('Кількість клієнтів') plt.show() Кількість клієнтів з/без голосової пошти 1200

df['voice mail plan'].value_counts().plot(kind='bar', color=['blue', 'orange'])

plt.title('Кількість клієнтів з/без голосової пошти')

2

2

False

True False

True

plt.ylabel('Середня оплата міжнародних дзвінків')

Name: total intl charge, dtype: float64

2.745050 2.750443

2.617019

3.221231

avg_intl_charge.unstack().plot(kind='bar', color=['blue', 'orange'])

2

In [4]:

Out[4]:

In [5]:

yes

avg_intl_charge

international plan churn

plt.xlabel('Ma∈ poymiнг')

0.0

In [6]:

200

150

100

0

seaborn

In [8]:

plt.show()

plt.show()

400

300

Частота 200

100

0

3

In [10]:

no

4

In [12]:

In [13]:

0

plt.xlabel(par1) plt.ylabel(par2)

0.0

In [14]:

б) міжнародними хвилинами і оплатою.

2.5

5.0

діаграмі розсіювання, і за значеням коєфіцієнту кореляції.

check_correlation(df, 'total intl minutes', 'total intl charge')

7.5

Залежність між цими двома змінними майже відсутня, що можна побачити, і на

plt.show()

50

def check_correlation(data, par1, par2): plt.figure(figsize=(10, 6))

sns.scatterplot(x=data[par1], y=data[par2])

corr_coeff = data[par1].corr(df[par2])

check_correlation(df, 'total intl minutes', 'total intl calls')

100

Викиди позначені на діаграмі розкиду поза вусами boxplot.

150

200

Хвилини в денний час

За допомогою діаграм розсіювання зробити висновки щодо залежності між: а) міжнародними хвилинами і дзвінками;

250

300

plt.figure(figsize=(9, 4))

400

500

600

Хвилини

Побудувати діаграму розмаху дзвінків в денний час загальну, визначити чи присутні викиди.

700

800

900

യായാ വ

350

 ∞

350

300

300

plt.xlabel('Хвилини') plt.ylabel('Частота') 400

plt.legend(title='Poymiнr', labels=['Maε', 'Hemaε'])

500

sns.histplot(df, x='total minutes', hue='international plan', bins=20) plt.title('Гістограма хвилин в залежності від підключеного роумінгу')

Побудувати гістограму хвилин в залежності від підключеного роумінгу.

600 Всі хвилини

Побудувати гістограму хвилин в нічний час

night_minutes = df['total night minutes']

plt.xlabel('Хвилини в нічний час')

plt.ylabel('Частота')

plt.hist(night_minutes, bins=20, color='blue') plt.title('Гістограма хвилин в нічний час')

Голосова пошта підключена

avg_intl_charge = df.groupby(['international plan', 'churn'])['total intl charge'].mean()

Кількість клієнтів 800 600 400 200

Голосова пошта підключена

б) мінімальну кількість хвилин в вечірній час для клієнтів з підключеною голосовою поштою і без;

min_evening_minutes = df.groupby('voice mail plan')['total eve minutes'].min()

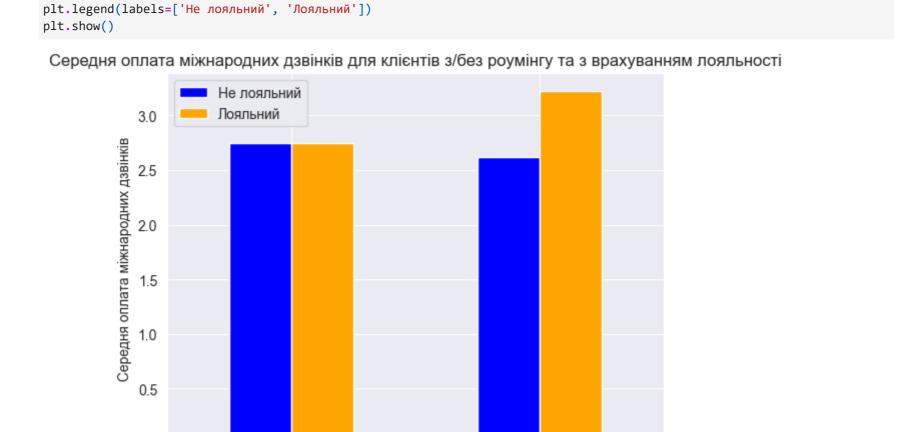
yes



в) середню оплату міжнародних дзвінків для клієнтів з підключеним роумінгом та без і з врахуванням лояльності.

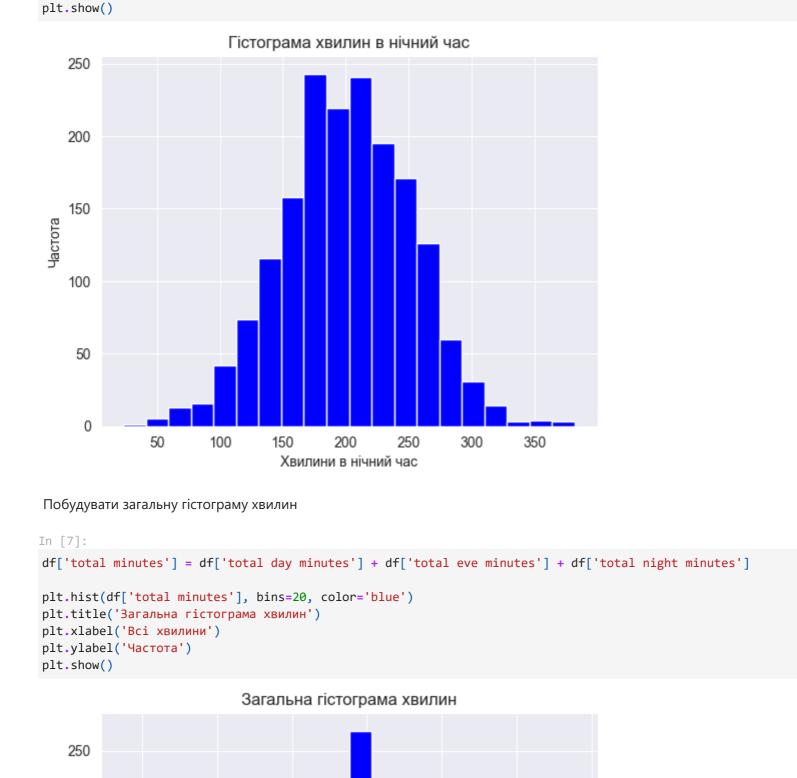
plt.title('Середня оплата міжнародних дзвінків для клієнтів з/без роумінгу та з врахуванням лояльності')

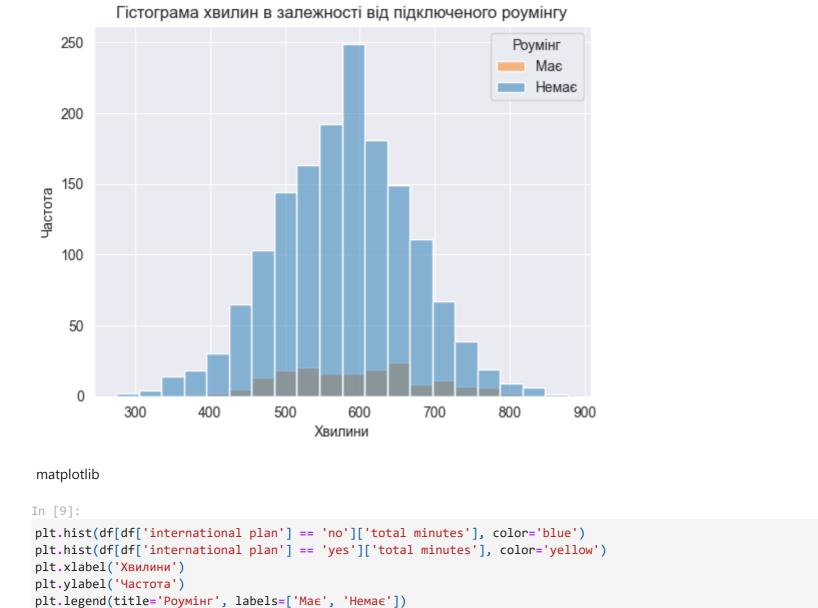
yes



Має роумінг

yes





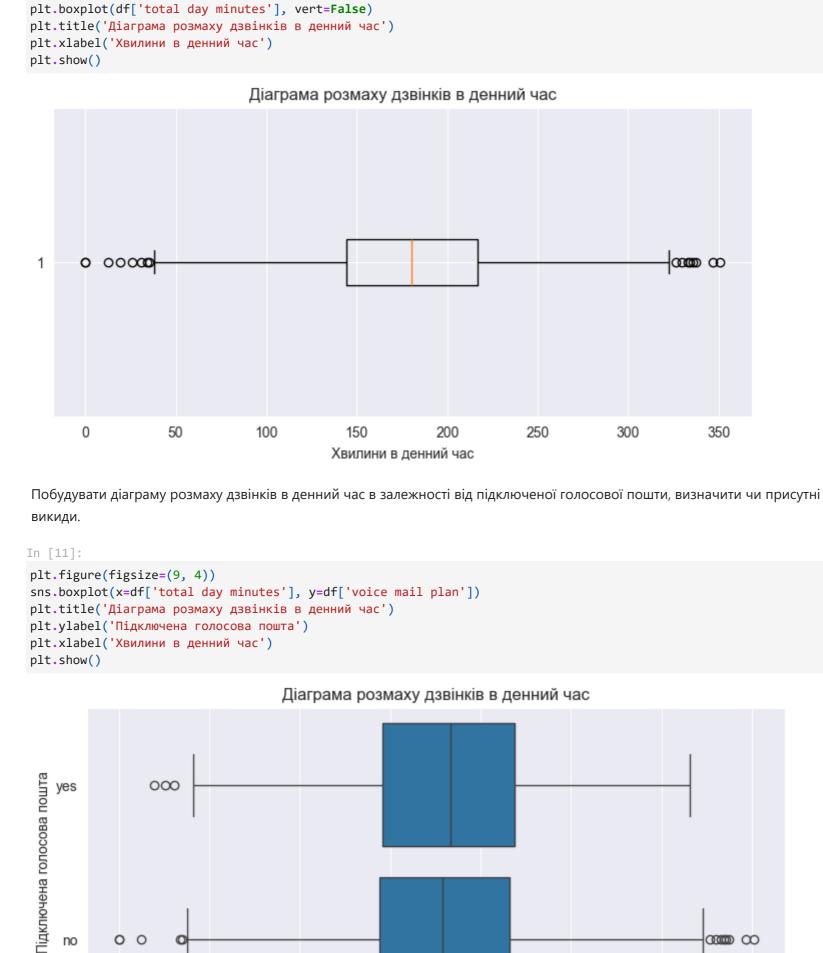
Роумінг Має

Немає

700

800

900



17.5 15.0

print(f"Коєфіцієнт кореляції між {par1} та {par2}: {round(corr_coeff, 4)}")

12.5 total intl calls 10.0 7.5 5.0 2.5 0.0

10.0

total intl minutes

12.5

15.0

17.5

20.0

