Lập trình phân tích dữ liệu 1 - kiểm tra giữa kỳ -đề 3

1

MSSV * □

21077161

2

Họ Tên * 🗔

Nguyễn Thanh Phới

Sẽ được xem xét

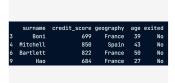
3

Data Manipulation with pandas

5 dòng dữ liệu đầu tiên của churn dataset như hình.

Đoạn code dưới sẽ cho ra kết quả nào churn[churn["credit_score"] < 600].head() ☐

| | surname | credit_score | geography | age | exited | |
|---|----------|--------------|-----------|-----|--------|--|
| 0 | Hargrave | 619 | France | 42 | Yes | |
| 1 | Hill | 608 | Spain | 41 | No | |
| 2 | Onio | 502 | France | 42 | Yes | |
| 3 | Boni | 699 | France | 39 | No | |
| 4 | Mitchell | 850 | Spain | 43 | No | |



| | surname | credit_score | geography | age | exited |
|----|---------|--------------|-----------|-----|--------|
| 7 | 0binna | 376 | Germany | 29 | Yes |
| 9 | Hao | 684 | France | 27 | No |
| 10 | Bearce | 528 | France | 31 | No |
| 11 | Andrews | 497 | Spain | 24 | No |
| 12 | Kay | 476 | France | 34 | No |

| | surname | credit_score | geography | age | exited |
|----|---------|--------------|-----------|-----|--------|
| 2 | Onio | 502 | France | 42 | Yes |
| 7 | 0binna | 376 | Germany | 29 | Yes |
| 8 | He | 501 | France | 44 | No |
| 10 | Bearce | 528 | France | 31 | No |
| 11 | Andrews | 497 | Spain | 24 | No |

(b

(c

Sẽ được xem xét

4

Data Manipulation with pandas

Đoạn code nào cho kết quả như hình

| | name | age |
|---|--------|-----|
| 0 | Onio | 32 |
| 1 | Bonnie | 42 |

Lập trình phân tích dữ liệu 1 - kiểm tra giữa kỳ -đề 3



Sẽ được xem xét

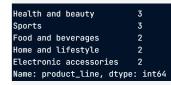
5

Data Manipulation with pandas

5 dòng dữ liệu đầu tiên của sales dataset như hình

Đoạn code dưới cho ra kết quả nào print(sales['product_line'].value_counts(sort = True).head())

| 5.99 | 3 | | |
|-------|------------------------|--------|-------|
| 25.28 | 1 | | |
| 10.23 | 1 | | |
| 12.99 | 1 | | |
| 6.99 | 1 | | |
| Name: | <pre>unit_price,</pre> | dtype: | int64 |





() a

b



Home and lifestyle

Sẽ được xem xét

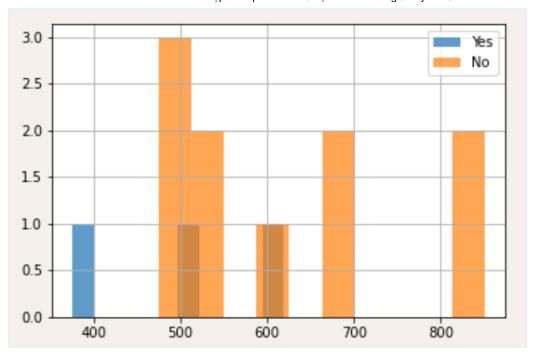
6

Data Manipulation with pandas

$$\label{lem:churn} \begin{split} & import\ matplotlib.pyplot\ as\ plt \\ & churn[churn["exited"]=="Yes"]["credit_score"].hist(alpha=0.7) \\ & churn[churn["exited"]=="No"]["credit_score"].hist(alpha=0.7) \end{split}$$

plt.show()

Đoạn code trên cho kết quả như hình, nhưng thiếu 1 dòng. Chọn đáp án đúng



- plt.legend("Yes","No")
- plt.legend(["Yes","No"])
- plt.legend(("Yes","No"))

Sẽ được xem xét

7

Introduction to Python

Hoàn thành đoạn code

x = [[1,2,3,4,5],["DataCamp","Python","Practice","Exercises"],[True, False, 0, 1]]print(x[?][:3])

để cho kết quả

['DataCamp', 'Python', 'Practice']

(a)

- O 2
- () -
- 3

☑ Sẽ được xem xét

8

x = [15,10,2,84] + [1,4,8,7,9]x.index(x.count(x[0]))

Kết quả của đoạn code trên là 🔲

- () 3
- () 2

Sẽ được xem xét

9

Introduction to Python

Dòng lệnh nào dùng để kiểm tra kiểu của số 5.0 🖫

- typeof 5.0
- Type(5.0)
- type(5.0)

Sẽ được xem xét

10

Introduction to Python

Cặp dòng lệnh nào cho cùng kết quả

(1)

- x[-3:] và x[1:]
- x[:] và x[:-1]
- x[1:] và x[:]

☑ Sẽ được xem xét

11

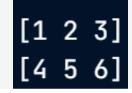
Introduction to Python and Numpy

Kết quả của đoạn code sau là gì 🖫

```
import numpy as np
np_array1 = np.array([1, 2, 3])
np_array2 = np.array([4, 5, 6])
np_array = np.array([np_array1, np_array2])

for val in np_array:
    print(val)
```

[1 2] [3 4]



b

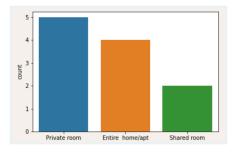
Sẽ được xem xét

12

Introduction to Data Visualization with Seaborn

room_type là list chứa các kiểu phòng khác nhau của Airbnb.

Doạn code sau cho kết quả như hình import seaborn as sns import matplotlib.pyplot as plt sns.___(x=room_type) plt.show()



() c

chọn method để hoàn thành đoạn code trên \Box

- O plot
- scatterplot
- countplot
- count

Sẽ được xem xét

13

Introduction to Statistics

data = [1, 1, 1, 3, 3, 3, 5]

Đoạn code nào cho kết quả là 3

p = np.mean(data)
print(p)

p = np.median(data)
print(p)

p = np.size(data)
print(p)

) a

() c

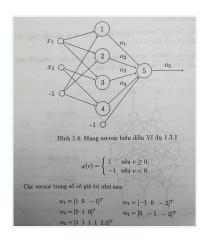
Sẽ được xem xét

14

neural network

Cho mạng neural 3 tầng với các vector trọng số và hàm kich hoạt như hình

Giá trị nào của x1 và x2 thì mạng neural này được kích hoạt



 $x_1=1 \text{ và } x_2=2$

1=3 và x2 = 1

x1=-2 và x2=0

x1=0 và x2=-1

☑ Sẽ được xem xét

15

Machine learning

Chọn câu sai 🗔

- Hiệu quả của mô hình nên được đánh giá sử dụng tập test
- Mô hình nên được huấn luyện trên cả tập train và tập test
- Trước khi huấn luyện mô hình, dữ liệu nên được chia thành tập train và tập test
- Hiệu quả của mô hình nên được đánh giá sau khi huấn luyện

Life Expectancy (years)

Sẽ được xem xét

16

Machine learning

Hoàn thành đoạn code để tạo đường hồi quy giữa tuổi thọ (life expectancy) và tỷ lệ sinh (fertility rate), cá dữ liệu này được lưu trữ tương ứng trong biến y và X. Predictions từ mô hình LinearRegression được vẽ như đường màu đỏ.

scatter; y

oplotl y

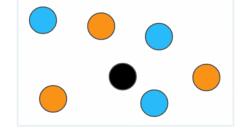
scatter; X

Sẽ được xem xét

17

Machine learning

Áp dụng mô hình KNN với k=3, thì điểm màu đen được phân loại vào nhóm màu xanh hay màu cam



Xanh

Cam

Cả 2 nhóm

Không thuộc nhóm nào cả

Sẽ được xem xét

18

Machine learning

Phát biểu nào sau đây không là ví dụ của ứng dụng máy học 🖽

- Dự báo bệnh tiểu đường dựa vào gluco trong máu và BMI
- Phân cụm khách hàng dựa vào hành vi mua sắm của họ trong quá khứ
- Dự báo điểm cuối kỳ dựa vào các điểm thành phần trước đó
- Tính toán năng lượng tiêu hao trung bình càu các loại xe hơi khác nhau

Sẽ được xem xét

19

Introduction to NumPy

Cho mảng numpy 2 chiều array_2D như hình

Dòng code nào dưới đây sẽ cho kết quả array([4., 6., 4., 9., 6., 3.])

array_2D.mean(axis=0)

- array_2D.mean()
- array_2D.cumsum(axis=1)

array([[5, 1, 6, 2, 4, 6], [5, 8, 0, 2, 1, 2], [2, 9, 6, 23, 13, 1]])

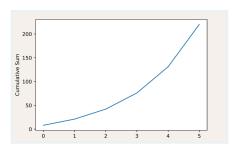
☑ Sẽ được xem xét

20

Hoàn thành đoạn code để vẽ đường tổng tổng tích lũy của màng arr

import matplotlib.pyplot as plt arr = np.array([8, 13, 21, 34, 55, 89]) plt.___(np.arange(6), arr.___()) plt.ylabel("Cumulative Sum")

- line; cumsum
- plot; sum
- scatter; sum
- plot; cumsum



Sẽ được xem xét

21

The productions DataFrame contains some missing values:

| | title | director | country |
|-----|--------------------|------------------|-----------------------|
| 33 | Squid Game | NaN | NaN |
| 37 | Angry Birds | NaN | Finland |
| 41 | Jaws | Steven Spielberg | United States |
| 127 | A Cinderella Story | Mark Rosman | United States, Canada |
| 361 | Charmed | NaN | United States |

chọn đoạn code để cho kết quả

1 Jaws 127 A Cinderella Story 573 Kung Fu Panda Name: title, dtype: object

cleaned = productions.dropna(subset=['director'])
print(cleaned['title'].head())

cleaned = productions.dropna(subset=['country']
print(cleaned['title'].head())

cleaned = productions.dropna()
print(cleaned['title'].head())

Tùy chọn 1

Tùy chọn 2

Tùy chọn 3

Sẽ được xem xét

22

Some columns in the productions DataFrame contain duplicates:

| | title | release_year country | |
|-----|-------------------------------|----------------------|--|
| 7 | The Great British Baking Show | 2021 UK | |
| 8 | The Great British Baking Show | 2021 UK | |
| 41 | Jaws | 1975 USA | |
| 127 | A Cinderella Story | 2004 USA | |
| 128 | A Cinderella Story | 2004 Canada | |
| 573 | Kung Fu Panda | 2008 USA | |
| 574 | Kung Fu Panda | 2008 China | |

df = productions[

productions.duplicated(
 subset = ['title','country'], keep = 'first')]
print(df.title)

Kết quả của đoạn code trên 🗔



7 The Great British Baking Show 8 The Great British Baking Show Name: title, dtype: object

8 The Great British Baking Show Name: title, dtype: object

а

(b

C

☑ Sẽ được xem xét

23

d = {'aa':30, 'bb':31, 'cc':32, 'dd':33}
? k, v in d.items():
 print(k, v)

hoàn thành đoạn code trên để cho kết quả

aa 30

bb 31

cc 32

dd 33

for each

| if |
|--|
| items |
| while |
| o for |
| |
| ☑ Sẽ được xem xét |
| 24 |
| <pre>d = {'a':0, 'b':1, 'c':2, 'd':3} if 'c' in ? : print(d['c']) else: print('c')</pre> |
| Hoàn thành đoạn code trên để cho kết quả |
| 2 🖳 |
| d.list() |
| d.items() |
| d.values() |
| (d |
| O d0 |
| |
| ☑ Sẽ được xem xét |
| 25 |
| d =? d[8] = 1 print(d) |
| Hoàn thành đoạn code trên để cho kết quả |
| {8: 1} □ ₀ |
| tuple() |
| dict[] |
| dict() |
| (list() |

| | | Sẽ | đu | ďC | xem | xét |
|--|--|----|----|----|-----|-----|
|--|--|----|----|----|-----|-----|

26

Decisions from p-values

The p-value, denoted here as p, is a measure of the amount of evidence to reject the null hypothesis or not. By comparing the p-value to the significance level, α , you can make a decision about which hypothesis to support.

Which of the following is the correct conclusion from the decision rule for a significance level lpha?

If the $p \leq lpha$, reject H_A .

If the $p \geq lpha$, reject H_A .

 \bigcirc ι

) b

If the $p \leq lpha$, don't reject H_0 .

If the $p \leq \alpha$, reject H_0 .

) (

d

Sẽ được xem xét

27

Which of the following statements about kNN model complexity is true?

- Increasing the value of k creates a more complex model
- Increasing the value of k, increases the risk of underfitting
- Increasing the value of k, increases the risk of overfitting

Sẽ được xem xét

28

Trong phân tích hồi quy tuyến tính, mục tiêu chính là gì? 🔲

Dự đoán giá trị của biến phụ thuộc dựa trên giá trị của biến độc lập

| Phân tích tương quan giữa hai biến độc lập |
|---|
| Xác định phương sai của biến phụ thuộc |
| Đánh giá sự ảnh hưởng của biến phụ thuộc lên biến độc lập |
| ☑ Sẽ được xem xét |
| 29 |
| Trong mô hình máy học KNN, việc dự đoán nhãn của một điểm dữ liệu mới được thực hiện như thế nào? 🔲 |
| Tính khoảng cách từ điểm dữ liệu mới đến tất cả các điểm dữ liệu trong tập huấn luyện và chọn ra K điểm gần nhất. |
| Sử dụng một phương pháp toán học phức tạp để dự đoán nhãn. |
| Lựa chọn ngẫu nhiên K điểm dữ liệu từ tập huấn luyện và tính trung bình nhãn của chúng. |
| Sử dụng một thuật toán tối ưu hóa để tìm ra điểm dữ liệu tốt nhất để dự đoán. |
| |
| E Sẽ được xem xét |
| 30 |
| Phương pháp EDA (Exploratory Data Analysis) trong phân tích dữ liệu nhằm mục đích gì? 🔲 |
| Xác định mối quan hệ nhân quả giữa các biến. |
| Khám phá và hiểu về cấu trúc và tính chất của dữ liệu. |
| Dự đoán giá trị của biến mục tiêu dựa trên biến đầu vào. |
| Tính toán các chỉ số thống kê mô tả cho dữ liệu. |
| |
| ☑ Sẽ được xem xét |
| 31 |
| Các công cụ nào thường được sử dụng trong EDA (Exploratory Data Analysis)? 🔲 |
| Regression và Classification. |
| regression va classification. |
| K-means clustering và Hierarchical clustering. |
| |

| Sẽ được xem xét |
|--|
| 32 |
| Phương pháp nào thường được sử dụng để xử lý giá trị ngoại lệ trong dữ liệu? 🔲 |
| Loại bỏ các quan sát có giá trị ngoại lệ khỏi tập dữ liệu. |
| Thay thế giá trị ngoại lệ bằng giá trị trung bình của biến. |
| Sử dụng kỹ thuật kiểm định t để xác định giá trị ngoại lệ. |
| Chuyển đổi giá trị ngoại lệ thành giá trị trung vị của biến. |
| ☑ Sẽ được xem xét |
| 33 |
| Đồ thị nào thường được sử dụng để biểu diễn mối quan hệ giữa hai biến liên tục? 🖽 |
| Biểu đồ cột (bar chart). |
| Biểu đồ hộp (box plot). |
| Biểu đồ dải (histogram). |
| Biểu đồ điểm (scatter plot). |
| ≅ Sẽ được xem xét |
| 34 |
| Trong NumPy, phương thức nào được sử dụng để tạo một mảng chứa các số ngẫu nhiên từ phân phối chuẩn (normal distribution)? 🖫 |
| numpy.random.rand() |
| numpy.random.randint() |
| numpy.random.randn() |
| numpy.random.uniform() |
| ≅ Sẽ được xem xét |
| 35 |
| Đại lượng nào trong thống kê mô tả cung cấp thông tin về phân bố của dữ liệu trên một khoảng giá trị? 口 |

| Phân vị |
|--|
| Độ lệch chuẩn |
| ○ Trung bình |
| O Phương sai |
| |
| ☑ Sẽ được xem xét |
| 36 |
| Trong machine learning, confusion matrix được sử dụng để làm gì? 🗔 |
| Đánh giá hiệu suất của mô hình phân loại. |
| Xác định sự biến động của dữ liệu. |
| Tính toán trọng số của các biến đầu vào. |
| Phân tích sự tương quan giữa các biến. |
| |
| EZ Sẽ được xem xét |
| 37 |
| Trong hồi quy, độ đo R-squared được sử dụng để đánh giá điều gì? 🔲 |
| |
| Độ chính xác của dự đoán so với dữ liệu thực tế. |
| Độ chính xác của dự đoán so với dữ liệu thực tế. Độ chính xác của mô hình hồi quy. |
| |
| Độ chính xác của mô hình hồi quy. |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. Độ lớn của hệ số hồi quy. |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. Độ lớn của hệ số hồi quy. Sế được xem xét |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. Độ lớn của hệ số hồi quy. Sẽ được xem xét 38 |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. Độ lớn của hệ số hồi quy. Sẽ được xem xét 38 Giá trị Z-score là gì trong phân phối chuẩn? |
| Độ chính xác của mô hình hồi quy. Mức độ tương quan giữa các biến độc lập và biến phụ thuộc. Độ lớn của hệ số hồi quy. Sẽ được xem xét 38 Giá trị Z-score là gì trong phân phối chuẩn? Sự chênh lệch giữa giá trị trung bình và giá trị trung vị của phân phối. |

| ☑ Sẽ được xem xét |
|---|
| 39 |
| Cross-validation giúp giảm thiểu vấn đề gì trong việc đánh giá hiệu suất của mô hình? 🔲 |
| Underfitting |
| Overfitting |
| Bias |
| Variance |
| |
| ☑ Sẽ được xem xét |
| 40 |
| Tứ phân vị thứ 25 (Q1) thường được hiểu như thế nào trong thống kê? 🔲 |
| Là giá trị trung bình của tập dữ liệu. |
| Là giá trị mà 25% dữ liệu nhỏ hơn và 75% dữ liệu lớn hơn trong tập dữ liệu đã sắp xếp. |
| Là giá trị mà 75% dữ liệu nhỏ hơn và 25% dữ liệu lớn hơn trong tập dữ liệu đã sắp xếp. |
| Là giá trị mà 50% dữ liệu nhỏ hơn và 50% dữ liệu lớn hơn trong tập dữ liệu đã sắp xếp. |
| |
| ☑ Sẽ được xem xét |
| 41 |
| Trong trường hợp nào sau đây, bạn nên sử dụng trung vị thay vì trung bình để mô tả trung tâm của dữ liệu? 🔲 |
| Khi dữ liệu có phân phối chuẩn. |
| Khi dữ liệu có outliers hoặc phân phối không đối xứng. |
| Chi dữ liệu có phân phối đều. |
| Khi dữ liệu có phân phối chuẩn và không có outliers. |
| |
| Sẽ được xem xét |
| 42 |
| Biểu đồ boxplot được sử dụng để biểu diễn như thế nào về phân phối của dữ liệu? 🔲 |
| Trung bình và phương sai. |

| \bigcirc | Phân phối chuẩn. |
|------------|------------------------|
| | Phân phối tương đối. |
| | Quartiles và outliers. |



Nội dung này được tạo bởi chủ sở hữu của biểu mẫu. Dữ liệu bạn gửi sẽ được gửi đến chủ sở hữu biểu mẫu. Microsoft không chịu trách nhiệm về quyền riêng tư hoặc thực tiễn bảo mật của khách hàng, bao gồm cả các biện pháp bảo mật của chủ sở hữu biểu mẫu này. Không bao giờ đưa ra mật khẩu của bạn.

Microsoft Forms | Các cuộc khảo sát, câu đố và cuộc thằm dò do Al cung cấp <u>Tạo biểu mẫu riêng của tôi</u> Quyền riêng tư và cookie | Điều khoản sử dụng