1. Phát biểu nào sau đây sai về vai trò của phân tích dữ liệu
   *  Hỗ trợ quá trình ra quyết định
   *  Dự báo điều gì sẽ xảy ra trong tương lai
   * ** Dự báo điều gì sẽ xảy ra trong tương lai**
   *  Phát hiện dữ liệu bất thường
2. Thuộc tính giới tính thuộc thang đo?
   * ** norminal**
   *  ordinal
   *  interval
   *  scale

**Correct**

1. Khi thuộc tính thu nhận nhận 1 trong ba giá trị: Cao, trung bình, thấp thì nó thuộc thang đo
   *  norminal
   * ** ordinal**
   *  interval
   *  ratio

**Correct**

1. Khi thu nhập nhận giá trị tính theo đơn vị tiền tệ, thì nó thuộc thang đo
   *  norminal
   *  ordinal
   *  interval
   * ** ratio**

**Correct**

1. Phát biểu nào sau đây sai?
   * ** thời tiết 30 độ sẽ nóng gấp đôi thời tiết 15 độ**
   *  Thu nhập 30 triệu/tháng gấp đôi thu nhập 15 triệu
   *  doanh thu 100 triệu trến tháng sẽ gấp 4 lần doanh thu 25 triệu/ tháng
   *  tình trạng hôn nhân là một thuộc tính dạng norminal

**Correct**

1. Điền từ còn thiếu vào phát biểu sau: một thuộc tính …. là thuộc tính nhận các giá theo cảm tính của người điền thông tin  
   * ****định lượng
   * ** định tính**
   *  cảm xúc
   *  interval
2. Thuộc tính … chỉ có các phép thống kê về tần số, tần suất mới có ý nghĩa, còn không thể thực hiện các phép tính trung bình
   * ** định lượng**
   *  định tính
   *  interval
   * ** ratio**
3. Thuộc tính thuộc thang đo … có thể thực hiện các phép tính tần số, tần suất, trung bình, trung vị, nhưng không thể đưa ra kết luận trên phép nhân, chia vì nó không có giá trị không tuyệt đối
   * ** interval**
   *  norminal
   *  ratio
   *  ordinal

**Correct**

1. Doanh thu là thuộc tính thuộc thang đo nào?
   *  interval
   *  norminal
   * ** ratio**
   *  ordinal

**Correct**

1. Công việc nào sau đây thuộc phân tích dữ liệu?
   *  thu thập dữ liệu
   *  làm sạch và biến đổi dữ liệu
   *  trực quan hóa dữ liệu
   * ** Tất cả các đáp án trên đều đúng**

**Correct**

[**[Quiz] Pandas – 1**](https://learn.codegym.vn/courses/phan-tich-du-lieu-voi-python/lessons/pandas-1/quizzes/quiz-pandas-1-2)

1. Đối tượng nào trong Pandas là một cấu trúc dữ liệu hai chiều có dạng bảng bao gồm các dòng và các cột được gắn nhãn và lập chỉ mục?
   * ** DataFrame**
   *  List
   *  Tuple
   *  Dictionary

**Correct**

1. Đối tượng nào trong Pandas là tập hợp các cặp khóa giá trị không có thứ tự, thường được sử dụng khi có một số lượng lớn dữ liệu
   *  DataFrame
   *  Tuple
   *  List
   * ** Dictionary**
2. Để thay thế giá trị ‘HN’ trong cột ‘Thanhpho’ thành giá trị ‘Ha noi’ ta thực hiện câu lệnh nào sau đây?
   * ** df['Thanhpho'].replace('HN','Ha noi',inplace = True)**
   *  df['Thanhpho'].replace('Ha noi','HN',inplace = True)
   *  df.replace([Thanhpho'='HN','Ha noi',inplace = True)
   *  df.replace([Thanhpho'='Ha noi','HN',inplace = True)

**Correct**

1. Để tạo DataFrame gồm một cột chứa 3 giá trị: Cao, Trung bình, Thấp ta có thể thực hiện câu lệnh nào sau đây?
   *  "lst = ['Cao','Trung bình','Thấp'] df = pd.DataFrame(lst)"
   *  "dic = {'Thu nhập':['Cao','Thấp','Trung bình']} df = pd.DataFrame(dic)"
   *  "tuples = [('Cao'),('Thấp'), ('Trung bình')] df = pd.DataFrame(tuples, columns=['Thu nhập'])"
   * ** Tất cả các đáp án trên đều đúng**

**Correct**

1. Để xác định số hàng, số cột của một DataFrame df ta thực hiện câu lệnh nào sau đây?
   * ** df.shape**
   *  df.shape[0]
   *  df.shape[1]
   *  df.columns

**Correct**

1. Trong Pandas object là kiểu dữ liệu gì?
   *  Không xác định
   *  Số nguyên
   * ** Chuỗi ký tự**
   *  Số thực

**Correct**

1. Trong Pandas để đọc dữ liệu vào DataFrame ta có thể sử dụng phương thức nào sau đây?
   * ** pd.read\_excel()**
   *  pd.read\_cvs()
   *  pd.read\_doc()
   *  pd.read\_son()
2. Để xóa cột ‘KhuyenMai’ và trả về thông tin của cột này ta thực hiện câu lệnh nào sau đây?
   *  del df['KhuyenMai']
   * ** df.pop('KhuyenMai')**
   *  df.drop('KhuyenMai', axis=1, inplace=True)
   *  Không thể trả về thông tin cột bị xóa
3. Để đổi tên cột ‘Price’ thành cột ‘DonGia’ ta thực hiện câu lệnh nào sau đây?
   *  df.rename(df['Price':'DonGia'],inplace=True)
   * ** df.rename(columns={'Price':'DonGia'},inplace=True)**
   *  df.rename(df['Price'='DonGia'],inplace=True)
   *  Không thể đổi tên cột thuộc tính

**Correct**

1. Để truy cập tới các phần tử có ‘MucUuTien’ khác 3 ta thực hiện câu lệnh nào sau đây?
   *  df.loc[df.MucUuTien >= 3]
   *  df.iloc[df.MucUuTien >= 3]
   * ** df.loc[df.MucUuTien != 3]**
   *  df.iloc[df.MucUuTien != 3]

**Correct**

[**[Quiz] Pandas – 2**](https://learn.codegym.vn/courses/phan-tich-du-lieu-voi-python/lessons/pandas-2-2/quizzes/quiz-pandas-2)

1. Để đo lường xu hướng tập trung của biến định lượng ta sử dụng ba giá trị nào sau đây?
   *  max, min, mode
   * ** mode, media, mean**
   *  mode, media, max
   *  mode, media, count

**Correct**

1. Để tính các giá trị thống kê của tất cả các biến định lượng trong DataFrame df, ta sử dụng phương thức nào sau đây?
   *  df.head
   *  df.mode()
   *  df.types()
   * ** df.describe()**

**Correct**

1. Phương thức mode() có ý nghĩa gì? Câu trả lời nào sau đây đúng nhất?
   *  Tìm giá trị xuất hiện nhiều nhất của một biến định lượng
   *  Tìm giá trị xuất hiện nhiều nhất của một biến định tính
   * ** Tìm giá trị xuất hiện nhiều nhất của một biến định tính hoặc biến định lượng**
   *  Chỉ dùng được cho biến định lượng

**Correct**

1. Xét biến định tính ‘df\_city’, câu lệnh df\_city.max() nhận kết quả nào sau đây?
   * ** Trả về một giá trị có độ dài lớn nhất của biến df\_city**
   *  Báo lỗi do max() chỉ thực hiện cho biến định lượng
   *  Trả về một giá trị xuất hiện nhiều nhất trong df\_city
   *  Tất cả các đáp án trên đều sai

**Correct**

1. Công cụ nào sau đây cho phép tổng hợp, lọc và phân tích dữ liệu?
   * ** pivot table**
   *  iloc
   *  query
   *  drop

**Correct**

1. Với mỗi ‘MatHang’ cho biết ‘SoLuong’ đã được bán lớn nhất theo từng ‘Thang’. Câu lệnh nào sau đây là chính xác?
   *  df.pivot\_table(values='SoLuong', index='MatHang', columns='Thang', aggfunc = 'max')
   *  df.pivot\_table(values='SoLuong', index='Thang', columns='MatHang', aggfunc = 'max')
   *  df.pivot\_table(values=['SoLuong'], index='MatHang', columns='Thang', aggfunc = {'max'})
   * ** Tất cả các đáp án trên đều đúng**
2. Để hợp nhất hai tập dữ liệu về Sinh viên trong đó mỗi tập dữ liệu mô tả những thuộc tính khác nhau. Ta sử dụng phương thức nào sau đây?
   *  merge
   * ** concat với axis=1**
   *  concat với axis=0
   *  Không thể hợp nhất do không có thuộc tính chung
3. Để hợp nhất hai tập dữ liệu có số lượng các bản ghi khác nhau mô tả các thông tin khác nhau về Sinh viên trong đó có chung thuộc tính MaSV. Ta sử dụng phương thức nào sau đây?
   *  merge
   *  concat với axis=1
   * ** Hai đáp án trên đều đúng**
   *  Không thể hợp nhất do không có cùng số lượng các bản ghi
4. Để hợp nhất các dòng dữ liệu tử các tập dữ liệu khác nhau, ta sử dụng phương thức nào sau đây?
   * ** concat và append**
   *  concat và merge
   *  merge và append
   *  Tất cả các đáp án trên đều đúng

**Correct**

1. Giả sử df1 chứa thông tin hóa đơn bán hàng tại một siêu thị của nhân viên A, và df2 chứa thông tin hóa đơn bán hàng của nhân viên B. Để hợp nhất thông tin bán hàng của hai nhân viên về 1 bản ghi duy nhất df, câu lệnh nào sau đây là chính xác?
   *  df = pd.concat([df1,df2], axis=1)
   * ** df = pd.concat([df1,df2], axis=0)**
   *  df = pd.append(df1,df2)
   *  df = df1.append(df2, axis=0)

**Correct**

[**[Quiz] Thao tác dữ liệu**](https://learn.codegym.vn/courses/phan-tich-du-lieu-voi-python/lessons/thao-tac-du-lieu-2/quizzes/quiz-thao-tac-du-lieu)

1. Lọc dữ liệu giúp thực hiện công việc gì sau đây?
   *  Xử lý giá trị khuyết thiếu
   *  Hiển thị các giá trị thống kê mô tả
   *  Tạo bộ dữ liệu mới từ tập dữ liệu đã có
   * ** Tạo bộ dữ liệu con từ tập dữ liệu đã có**
2. Phát biểu nào sau đây không chính xác?
   *  Ta có thể lọc dữ liệu theo hàng, theo cột
   *  Ta có thể lọc dữ liệu theo điều kiện
   * ** Phương thức lọc query chạy chậm hơn phương thức lọc filter**
   *  Ta có thể lọc theo tên cột và chỉ số dòng

**Correct**

1. Thuộc tính filter cho phép?
   * ** Lọc các bản ghi rời rạc, các chỉ số được ngăn cách nhau bởi dấu phẩy**
   *  Lọc các bản ghi liên tiếp, chỉ số đầu ngăn cách chỉ số cuối bởi dấu :
   *  Các tham số là các biểu thức điều kiện đơn
   *  Các tham số là các biểu thức điều kiện ghép
2. Lệnh nhóm dữ liệu groupby thực hiện theo thứ tự nào sau đây?
   * ** “(1) Chia dữ liệu thành các nhóm theo tiêu chí.  
     (2) Áp dụng các phép tính cho từng nhóm một cách độc lập.  
     (3) Kết hợp các kết quả từ mỗi nhóm vào một cấu trúc dữ liệu.”**
   *  (1) Áp dụng các phép tính cho từng nhóm một cách độc lập.  
     (2) Chia dữ liệu thành các nhóm theo tiêu chí.  
     (3) Kết hợp các kết quả từ mỗi nhóm vào một cấu trúc dữ liệu.
   *  (1) Chia dữ liệu thành các nhóm theo tiêu chí.  
     (2) Kết hợp các kết quả từ mỗi nhóm vào một cấu trúc dữ liệu.  
     (3) Áp dụng các phép tính cho từng nhóm một cách độc lập.
   *  (1) Kết hợp các kết quả từ mỗi nhóm vào một cấu trúc dữ liệu.  
     (2) Chia dữ liệu thành các nhóm theo tiêu chí.  
     (3) Áp dụng các phép tính cho từng nhóm một cách độc lập.
3. Câu lệnh nào sau đây thực hiện được yêu cầu “Với mỗi chuyên nganh ‘ChuyenNganh’ hãy tính trung bình cộng số nguyện vọng ‘SoNguyenVong’ đã đăng ký”?
   *  df.groupby.mean('ChuyenNganh')['SoNguyenVong']
   * ** df.groupby('ChuyenNganh')['SoNguyenVong'].mean()**
   *  df.groupby.mean('SoNguyenVong')['ChuyenNganh']
   *  df.groupby('SoNguyenVong')['ChuyenNganh'].mean()

**Correct**

1. Phát biểu nào sau đây là chính xác?
   *  groupby chỉ cho phép nhóm dữ liệu theo một tiêu chí nhưng cho phép có nhiều tiêu chí tính toán.
   *  groupby cho phép nhóm dữ liệu theo nhiều tiêu chí nhưng chỉ cho phép một tiêu chí tính toán.
   *  groupby chỉ cho phép nhóm dữ liệu và tính toán theo một tiêu chí đơn.
   * ** Tất cả các đáp án trên đều sai**

**Correct**

1. Trong pandas để thêm cột dữ liệu mới vào một dataframe ta thực hiện câu lệnh nào sau đây?
   *  Có thể hợp nhất các cột từ nhiều dataframe.
   *  Có thể ghép thông tin từ các cột dữ liệu có sẵn.
   *  Có thể tách thông tin từ các cột dữ liệu có sẵn.
   * ** Tất cả các đáp án trên đều đúng.**

**Correct**

1. Thêm cột ‘XepLoai’ có giá trị ‘Dat’ nếu ‘ThuNhap’ >= 20, ngược lại nhận giá trị ‘KhongDat’. Câu lệnh nào sau đây chính xác?
   *  df['XepLoai'] = np.select(df['ThuNhap'] >= 20,'Dat','KhongDat' )
   *  df['XepLoai'] = np.select(df['ThuNhap'] >= 20,'KhongDat','Dat' )
   * ** df['XepLoai'] = np.where(df['ThuNhap'] >= 20,'Dat','KhongDat' )**
   *  df['XepLoai'] = np.where(df['ThuNhap'] >= 20,'KhongDat','Dat' )

**Correct**

1. Hàm tg.astype(str) có ý nghĩa nào sau đây?
   *  Đổi str về kiểu chuỗi
   * ** Đổi tg về kiểu chuỗi**
   *  Đổi str về kiểu tg
   *  Đổi tg về kiểu str
2. Để lấy thông tin tháng lập báo cáo (có dạng số) từ thuộc tính ‘data\_report’ có dạng ’15/08/2021′ câu lệnh nào sau đây là chính xác?
   *  df['month']=df['data\_report'].str.split('/').str[1]
   *  df['month']=pd.to\_datetime(df['date\_report'],format='%Y-%m-%d').dt.month
   * ** df['month']=pd.to\_datetime(df['date\_report'],format='%d/%m/%Y').dt.month**
   *  Tất cả các đáp án trên đều đúng

**Bài thi kết thúc khóa Phân tích dữ liệu với Python**

1. Đối tượng nào trong Pandas là một cấu trúc dữ liệu hai chiều có dạng bảng bao gồm các dòng và các cột được gắn nhãn và lập chỉ mục?
   * ** DataFrame**
   *  List
   *  Tuple
   *  Dictionary

**Correct**

1. Thuộc tính thuộc thang đo … có thể thực hiện các phép tính tần số,tần suất, trung bình, trung vị, nhưng không thể đưa ra kết luận trên phép nhân, chia vì nó không có giá trị không tuyệt đối
   * ** interval**
   *  norminal
   *  ratio
   *  ordinal

**Correct**

1. Hàm tg.astype(str) có ý nghĩa nào sau đây?
   *  Đổi str về kiểu chuỗi
   * ** Đổi tg về kiểu chuỗi**
   *  Đổi str về kiểu tg
   *  Đổi tg về kiểu str

**Correct**

1. Trong Pandas, để hợp nhất các dòng dữ liệu tử các tập dữ liệu khác nhau, ta sử dụng phương thức nào sau đây?
   * ** concat và append**
   *  concat và merge
   *  merge và append
   *  Tất cả các đáp án trên đều đúng

**Correct**

1. Một Figure có thể vẽ được tối đa bao nhiêu loại biểu đồ?
   *  2
   *  4
   *  1
   * ** Không giới hạn**

**Correct**

1. Để tính các giá trị thống kê của tất cả các biến định lượng trong DataFrame df, ta sử dụng phương thức nào sau đây?
   *  df.head
   *  df.mode()
   *  df.types()
   * ** df.describe()**

**Correct**

1. Phương pháp mã hóa nào phù hợp với dữ liệu là mức độ hài lòng của khách hàng: kém, tốt, rất tốt, tuyệt vời?
   *  Label encoding
   *  Ordinal encoding
   *  One-hot encoding
   * ** Frequency Encoding**
2. Hàm nào trong seaborn có thể dùng để load dữ liệu có sẵn?
   * ** load\_dataset**
   *  load\_csv
   *  load
   *  read\_dataset

**Correct**

1. Dữ liệu bị khuyết kiểu datetime hiển thị trong pandas là gì?
   * **** NaN
   *  NA
   *  None
   * ** NaT**
2. Công cụ nào sau đây cho phép tổng hợp, lọc và phân tích dữ liệu?
   * ** pivot table**
   *  iloc
   *  query
   *  drop

**Correct**

1. Đâu là dữ liệu phân loại thứ tự (Có thể chọn nhiều đáp án)?
   * ** Cỡ giày**
   *  Tên Hãng thời trang
   *  Tên quốc gia
   * ** Trình độ học vấn**

**Correct**

1. Biểu đồ thể hiện xu hướng của dữ liệu thay đổi theo thời gian tốt nhất là biểu đồ nào?
   *  Biểu đồ tròn
   *  Biểu đồ Scatter
   *  Biểu đồ Histogram
   * ** Biểu đồ đường**

**Correct**

1. Biểu đồ thể hiện mối quan hệ giữa các thuộc tính là?
   *  Histogram
   * ** Biểu đồ đường**
   * ** Biểu đồ Scatter**
   *  Biểu đồ cột
2. Biểu đồ nào thể hiện phân bố của dữ liệu?
   * ** Histogram**
   *  Đường
   *  Hộp
   *  Thác nước

**Correct**

1. s = pd.Series([‘red’, ‘red’, ‘yellow’, ‘green’, ‘yellow’]). Cho 1 Series như trên, hàm nào sau đây thực hiện label encoding?
   * ** s.astype('category').cat.codes**
   *  s.cat.codes
   *  s.astype('int64').cat.codes
   *  s.astype('category').codes

**Correct**

1. Kiểu dữ liệu nào sau đây nên mã hóa bằng one-hot encoding?
   *  Cỡ giày
   * ** Tên Hãng thời trang**
   * ** Tên quốc gia**
   *  Trình độ học vấn
2. Kiểm định pearson được dùng để làm gì?
   * ** Kiểm định tương quan tuyến tính giữa hai thuộc tính định lượng**
   *  Kiểm định tương quan tuyến tính giữa hai thuộc tính định tính
   *  Kiểm định tương quan tuyến tính giữa hai thuộc tính thứ bậc
   *  Phát hiện dữ liệu bất thường
3. Điền từ còn thiếu vào phát biểu sau: “Một thuộc tính …. là thuộc tính nhận các giá theo cảm tính của người điền thông tin”.
   *  định lượng
   * ** định tính**
   *  cảm xúc
   *  interval

**Correct**

1. Doanh thu là thuộc tính thuộc thang đo nào?
   *  interval
   *  norminal
   * ** ratio**
   *  ordinal

**Correct**

1. Dùng loại biểu đồ nào có thể giúp ta phát hiện ngoại lai?
   * ** Box Plot**
   *  Bar Plots
   *  Pie Plots
   *  Line Plot

**Correct**

1. Khi thuộc tính thu nhận nhận 1 trong ba giá trị: Cao, trung bình, thấp thì nó thuộc thang đo nào?
   *  norminal
   * ** ordinal**
   *  interval
   *  ratio

**Correct**

1. Để xóa cột ‘KhuyenMai’ và trả về thông tin của cột này ta thực hiện câu lệnh nào sau đây?
   *  del df['KhuyenMai']
   * ** df.pop('KhuyenMai')**
   *  df.drop('KhuyenMai', axis=1, inplace=True)
   *  Không thể trả về thông tin cột bị xóa

**Correct**

1. Kiểm định giả thuyết giúp:
   * ** Kiểm chứng thông tin**
   *  Trình bày dữ liệu
   *  Dự báo điều gì sẽ xảy ra trong tương lai
   *  Phát hiện dữ liệu bất thường
   *  Xác suất mắc sai lầm

**Correct**

1. Ưu điểm của seaborn với matplotlib là gì?
   *  Cộng đồng lớn hơn
   *  Nhanh hơn
   *  Tinh chỉnh được nhiều tham số hơn
   * ** API đơn giản hơn**

**Correct**

1. Sai lầm loại I là ?
   * ** Bác bỏ một giả thuyết đúng**
   *  Chấp nhận một giả thuyết sai
   *  viết sai giả thuyết kiểm định
   *  sử dụng nhầm phương pháp kiểm định
2. Thuộc tính giới tính thuộc thang đo nào?
   * ** norminal**
   *  ordinal
   *  interval
   *  scale

**Correct**

1. Để đo lường xu hướng tập trung của biến định lượng ta sử dụng ba giá trị nào sau đây?
   *  max, min, mode
   * ** mode, media, mean**
   *  mode, media, max
   *  mode, media, count

**Correct**

1. Để thay thế giá trị ‘HN’ trong cột ‘Thanhpho’ thành giá trị ‘Ha noi’ ta thực hiện câu lệnh nào sau đây?
   * ** df['Thanhpho'].replace('HN','Ha noi',inplace = True)**
   *  df['Thanhpho'].replace('Ha noi','HN',inplace = True)
   *  df.replace([Thanhpho'='HN','Ha noi',inplace = True)
   *  df.replace([Thanhpho'='Ha noi','HN',inplace = True)

**Correct**

1. Phát biểu nào sau đây là chính xác?
   *  groupby chỉ cho phép nhóm dữ liệu theo một tiêu chí nhưng cho phép có nhiều tiêu chí tính toán.
   *  groupby cho phép nhóm dữ liệu theo nhiều tiêu chí nhưng chỉ cho phép một tiêu chí tính toán.
   *  groupby chỉ cho phép nhóm dữ liệu và tính toán theo một tiêu chí đơn.
   * ** Tất cả các đáp án trên đều sai**

**Correct**

1. Trong kiểm định Chi bình phương nếu pvalue = 0.8 thì có nghĩa là?
   *  Tương quan phi tuyến mạnh
   *  Tương quan tuyến tính nghịch
   * ** Không có tương quan**
   *  Tương quan tuyến tính thuận