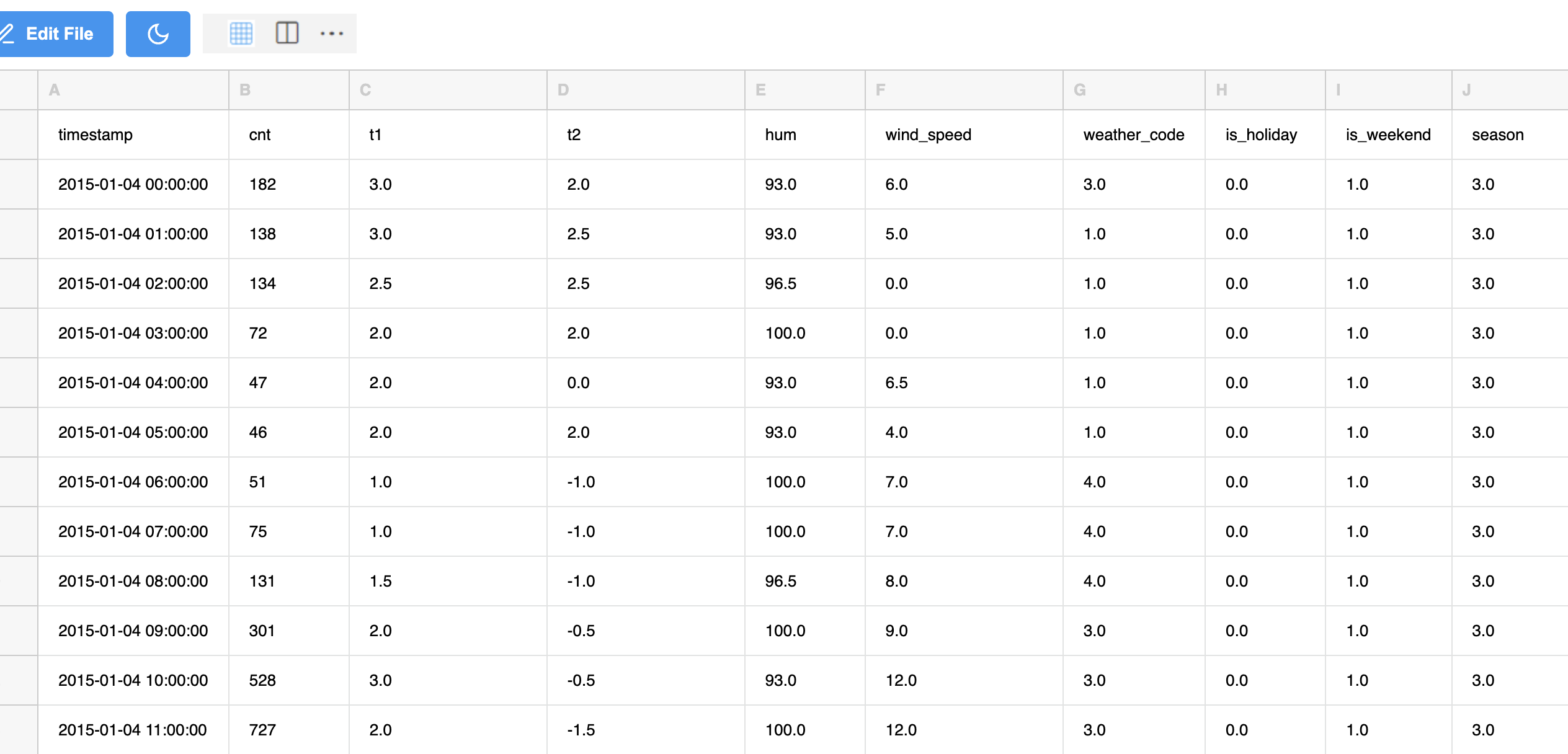
**Phân tích và diễn giải Nhu cầu sử dụng xe đạp & Hành vi của khách hàng**

Từ bộ dữ liệu được khách hàng cung cấp (Có 17.714 dòng dữ liệu với 10 cột thông tin như bảng dưới) : Bạn hãy thử phân tích các yêu tố ảnh hưởng đến nhu cầu sử dụng xe đạp của khách hàng. *(Hình ảnh dữ liệu mẫu)*



**Ý nghĩa từng cột dữ liệu:**

|  |  |
| --- | --- |
| * Timestamp: Thời gian sử dụng | * wind\_speed: Tốc độ gió Km/h |
| * cnt: Số lượt đạp xe | * is\_holiday: Ngày lễ *(1 ngày lễ / 0 không nghỉ lễ)* |
| * t1: Nhiệt độ thực tế C | * is\_weekend: Ngày cuốt tuần (1 nếu ngày là cuối tuần) |
| * t2: Nhiệt độ cảm giác | * Season: Mùa (0: mùa xuân; 1: mùa hè; 2: mùa thu; 3: mùa đông). |
| * hum: % Độ ẩm | * weather\_code: Mã thời tiết |

Trong đó cột: “**weather\_code”** Chi tiết tương tứng với từng mã dữ liệu như sau:

* + 1 = Rõ ràng ; hầu như rõ ràng nhưng có một số giá trị với sương mù / sương mù / các mảng sương mù / sương mù trong vùng lân cận
  + 2 = mây rải rác / mây ít
  + 3 = Những đám mây vỡ
  + 4 = Nhiều mây
  + 7 = Mưa/ Mưa nhẹ/ Mưa nhẹ
  + 10 = mưa với giông bão
  + 26 = tuyết rơi
  + 94 = Sương mù đóng băng

**Bước 1:**

* + - Cài đặt các thư viên cần thiết
    - Tải tập dữ liệu mẫu và hiển thị dữ liệu kiểm tra
    - Nhận số liệu thống kê cơ bản.

Bước 2: EDA

-

1. **Nhìn nhận mối tương quan giữa các đặc trưng của dữ liệu**

A screenshot of a graph

AI-generated content may be incorrect.

🌡️ Mối quan hệ giữa “Nhiệt độ cảm giác” và “Nhiệt độ thực tế”

📊 Nhìn vào biểu đồ chúng ta thấy “Nhiệt độ cảm giác” vs “Nhiệt độ thực tế” có mối tương quan rất mạnh = 0.99, ngần như là mỗi quan hệ tuyến tính. Điều này cũng dễ hiểu vì nhiệt độ cảm giác thường chịu ảnh hưởng bới nhiệt độ thực tế.

📉 Khi mối **tương quan theo cặp** của tất cả các biến (Đặc trưng) khác được kiểm tra, **các giá trị tương quan thấp**đã được phát hiện.

A screenshot of a graph

AI-generated content may be incorrect.

Ở đây chúng ta xem 🎯Biến mục tiêu là “Số lượt sử dụng” Xe đạp.

Nhìn vào biểu đồ phân phối của "Số lượt sử dụng", chúng ta thấy những điều sau:

*Hình dạng phân phối*: Phân phối này rõ ràng là lệch phải (hoặc lệch dương). Đỉnh của biểu đồ (nơi có tần suất cao nhất) nằm ở các giá trị thấp của "Số lượt sử dụng", và đuôi của biểu đồ kéo dài về phía bên phải, cho thấy có một số lượng nhỏ các chuyến đi có số lượt rất cao.

*Giá trị xuất hiện thường xuyên nhất (đỉnh):* Đỉnh của biểu đồ nằm ở khoảng giá trị rất thấp, gần như bằng 0. Điều này cho thấy phần lớn các quan sát có số lượt đi thấp.

*Phạm vi giá trị:* Biểu đồ trải dài từ khoảng 0 đến hơn 7500 (dựa trên trục hoành). Điều này cho thấy sự biến động lớn trong số lượt đi.

*Sự tập trung của dữ liệu:* Dữ liệu tập trung chủ yếu ở các giá trị thấp của "Số lượt đi". Khi số lượt đi tăng lên, tần suất xuất hiện của các giá trị này giảm đi nhanh chóng.

*Khoảng trống và đỉnh phụ:* Tôi không thấy có khoảng trống lớn nào trong phân phối. Có vẻ như có một "vai" nhỏ hoặc một đỉnh thứ hai không rõ ràng lắm ở khoảng giá trị thấp hơn 2500, nhưng đỉnh chính vẫn ở gần 0.

Điều này có thể gợi ý một vài điều:

Có thể có các sự kiện đặc biệt, ngày lễ, hoặc điều kiện thời tiết thuận lợi làm tăng đáng kể số lượt đi.

Có thể có sự khác biệt lớn về nhu cầu di chuyển giữa các thời điểm khác nhau.

Hiểu được sự phân phối của biến mục tiêu là bước quan trọng đầu tiên trong phân tích dữ liệu. Nó giúp chúng ta hình dung được đặc điểm của biến mà chúng ta muốn dự đoán hoặc giải thích, và có thể ảnh hưởng đến việc lựa chọn các phương pháp phân tích và mô hình phù hợp sau này.

**📊 Kết quả**

* "Trực quan hóa nhu cầu sử dụng xe đạp" kiểm tra hiệu quả các yếu tố ảnh hưởng đến nhu cầu sử dụng xe đạp bằng cách sử dụng các kỹ thuật trực quan hóa và phân tích dữ liệu. Những phát hiện chính từ Bài toán bao gồm:

**🌱 Các mẫu theo mùa và tạm thời:**

* Phân tích cho thấy xu hướng thời gian rõ ràng trong nhu cầu xe đạp, với mức sử dụng cao hơn được quan sát thấy trong một số tháng và mùa nhất định. Việc sử dụng xe đạp mạnh mẽ hơn trong những tháng mùa hè, phù hợp với điều kiện thời tiết thuận lợi.
* Các mô hình nhu cầu sử dụng xe ngày trong tuần và cuối tuần khác nhau, phản ánh các ngày làm việc trong tuần và các hoạt động giải trí cuối tuần.

**🌦 Tác động của điều kiện thời tiết:**

* Điều kiện thời tiết, đặc biệt là nhiệt độ, độ ẩm và lượng mưa, ảnh hưởng đáng kể đến việc sử dụng xe đạp. Trong khi thời tiết ấm hơn có mối tương quan tích cực với nhu cầu cao hơn, các điều kiện thời tiết bất lợi như mưa lớn dẫn đến giảm đáng kể mức sử dụng.

**🚲 Phân tích hành vi của người dùng:**

* Các mô hình sử dụng giữa những người dùng khác nhau. Người dùng đóng góp đáng kể vào nhu cầu vào các ngày trong tuần, có thể là do đi lại hàng ngày của họ. Ngược lại, người dùng có xu hướng ít hoạt động hơn vào cuối tuần và ngày lễ.

**🎉 Tác động của các ngày lễ và sự kiện đặc biệt:**

* Dữ liệu cho thấy việc sử dụng xe đạp giảm trong các ngày lễ và cuối tuần. Điều này có thể là do nhu cầu vận chuyển giảm hoặc các hoạt động khác nhau trong các ngày lễ.

**📈 Tầm quan trọng của việc trực quan hóa dữ liệu:**

* Bài toán nhấn mạnh tầm quan trọng của việc trực quan hóa trong việc xác định xu hướng và trích xuất những hiểu biết có thể hành động. Biểu đồ, bản đồ nhiệt và bảng điều khiển tương tác cung cấp sự hiểu biết rõ ràng về dữ liệu phức tạp.

**💡 Các đề xuất**

**🛠 Tối ưu hóa phân bổ tài nguyên:**

* Dựa trên xu hướng thời gian và thời tiết, các dịch vụ sử dụng xe đạp có thể tối ưu hóa việc phân bổ tài nguyên; ví dụ: định vị lại xe đạp và lên lịch bảo trì trong thời gian thấp điểm.

**🌧 Các chiến lược để thích ứng với điều kiện thời tiết:**

* Cung cấp các chương trình khuyến mãi hoặc ưu đãi trong điều kiện thời tiết bất lợi có thể giúp duy trì nhu cầu. Ngoài ra, việc cung cấp xe đạp chịu được thời tiết và thiết bị bảo hộ có thể làm tăng sự hài lòng của người dùng.

**🔄 Chuyển đổi người dùng thành người dùng đã đăng ký:**

* Các chiến lược như giảm giá cho tư cách thành viên hoặc tiếp thị được nhắm mục tiêu có thể làm tăng sự hài lòng của người dùng với các dịch vụ sử dụng xe đạp.

**🏗 Quy hoạch cơ sở hạ tầng:**

* Xác định các khu vực có nhu cầu cao và thời gian cao điểm có thể hướng dẫn các quyết định làm việc cơ sở hạ tầng, đảm bảo xe đạp có sẵn ở đâu và khi chúng có nhu cầu cao nhất.