

MÔ TẢ DATASETS

1. Executive Sales Report

Mô tả tình huống: Bạn là một nhà phân tích làm việc tại một nhà bán lẻ lớn. Giám đốc bán hàng yêu cầu bạn chuẩn bị một báo cáo mà anh ta có thể gửi cho các giám đốc điều hành về hiệu suất bán hàng và để đảm bảo rằng nó có vẻ tốt.

Mô tả dataset: 5 tables:

- Sales table: Đơn hàng theo ngày, kênh bán hàng, team bán hàng, cửa hàng và mặt hàng
- Regions: Khu vực
- Sales Team: Team bán hàng
- Products: Danh sách sản phẩm
- Store Locations: Thông tin về vị trí địa lý của cửa hàng

2. COVID-19 Reporting

Mô tả tình huống: Tình hình dịch bệnh covid thế giới và VN. Học viên cần tìm các nguồn để thu thập dữ liệu.

Tham khảo: <https://vnexpress.net/covid-19/covid-19-viet-nam>

3. Environmental Data Reporting

Mô tả tình huống: Bạn là nhà phân tích dữ liệu và chuyên gia trực quan hóa tại một tổ chức môi trường lớn ở Hoa Kỳ khu vực Trung Đại Tây Dương đại diện cho Delaware, Columbia/Washington DC, Maryland, Pennsylvania, New York, New Jersey, Virginia và Tây Virginia. Bạn được giao nhiệm vụ phát triển một báo cáo để truyền đạt dữ liệu môi trường có liên quan giúp hình dung các mô hình đang phát triển. Các chỉ số đã được chọn sẵn và dữ liệu đã được thu thập. Công việc của bạn là xác định cách tốt nhất để trực quan hóa dữ liệu và phát triển báo cáo để truyền đạt thông tin chính.

Các chỉ số:

Một chỉ số là một giá trị số được lấy từ các phép đo thực tế của một tác nhân gây căng thẳng, trạng thái, mức độ phơi nhiễm hoặc sức khỏe con người hoặc tình trạng sinh thái trên một khu vực địa lý cụ thể, mà xu hướng theo thời gian thể hiện hoặc thu hút sự chú ý đến các xu hướng cơ bản trong điều kiện môi trường. Các chỉ số là độc lập và với mục đích trực quan, không có mối tương quan hoặc nhân quả nào được kết nối hoặc ám chỉ đến.

Báo cáo cần bao gồm thông tin về tất cả bốn chỉ số môi trường quan trọng đối với khu vực:

- Lượng mưa lớn: tổng hợp hàng năm về lượng mưa hàng ngày từ 23 địa điểm (22 địa điểm sân bay, 1 địa điểm phi sân bay) trong toàn khu vực.

- Chất độc không khí xung quanh: nồng độ chất độc trong không khí có hại cho sức khỏe con người và sinh thái tại các điểm trong vùng. Lưu ý rằng nồng độ được biểu thị bằng $\mu\text{g} / \text{m}^3$ (microgam trên mét khối không khí).
- Tỷ lệ mắc bệnh hen suyễn: tỷ lệ mắc bệnh hen suyễn ở người lớn (thông qua mô hình hóa) theo cấp điều tra dân số và tỷ lệ hiện mắc bệnh ở trẻ em theo tỷ lệ phần trăm của tiểu bang trong toàn khu vực. Lưu ý rằng trong dữ liệu về bệnh hen suyễn của trẻ em, "bị ức chế" có nghĩa là con số quá thấp để hiển thị mà không có khả năng tiết lộ danh tính của ai đó và "không ổn định" (thường trùng lặp với "bị kiềm chế") có nghĩa là các con số được sử dụng để tính toán tỷ lệ thấp đến mức chúng không đáng tin cậy về mặt thống kê.
- Tiếp xúc với Ô nhiễm của Con người: các địa điểm bị ô nhiễm trong khu vực có mối quan tâm về phơi nhiễm của con người. Chỉ có ba loại trang web: Tiềm năng cần quan tâm, Không quan tâm đến việc phơi bày và không đủ dữ liệu.

Người sử dụng:

- Các nhà khoa học dữ liệu khu vực sẽ sử dụng báo cáo làm điểm khởi đầu để đánh giá toàn diện hơn về các xu hướng môi trường
- Người quản lý cấp trên không quen thuộc với chi tiết của bộ dữ liệu, nhưng sẽ sử dụng báo cáo để theo dõi các xu hướng môi trường chính và sẽ sử dụng các hình ảnh trực quan được tạo ra để truyền đạt thông tin này trong các bài thuyết trình bên ngoài
- Các nhà hoạch định chính sách, những người cần hiểu nhanh dữ liệu để hỗ trợ các quyết định về cách phân bổ vốn cho các khía cạnh của khu vực mà họ cần chú ý nhất

4. Insurance Complaints Data

Mô tả tình huống: Bạn đang làm việc tại một công ty bảo hiểm lớn sắp trải qua cuộc kiểm toán nội bộ liên quan đến thủ tục khiếu nại. Công ty đang có tư duy tiến bộ và hiện có một kho dữ liệu cùng với một số mô hình. Ban quản lý đã yêu cầu bạn xem xét mô hình và cung cấp thông tin chi tiết cho cấp quản lý để có thể làm nổi bật bất kỳ vấn đề nào và chuẩn bị chúng cho cuộc đánh giá. Bạn phải tự lựa chọn các vấn đề phân tích mà bạn cho là quan trọng

Một số yếu tố cần xem xét:

- Dữ liệu 2 năm gần đây nhất là điều mà các bên liên quan quan tâm nhất
- Khả năng xem các thay đổi trạng thái và thời điểm chúng xảy ra
- Khiếu nại được chia nhỏ theo các dimensions
- Sự hài lòng của khách hàng
- Brokers vi phạm lỗi nghiêm trọng nhất

Mô tả dataset: 12 bảng dữ liệu, bao gồm 2 bảng fact Complains Data và Status History Data, ngoài ra có các bảng dimensions để mô tả và làm rõ dữ liệu

5. Transport & Shipping Data

Mô tả tình huống: Đây là tập dữ liệu của một khách hàng đang muốn sử dụng Power BI cho các mục đích báo cáo và phân tích. Họ đã lấy một bản trích xuất từ hệ thống nguồn và đưa nó vào excel. Trong các trường hợp bình thường, nhóm sẽ làm việc với excel này và tạo ra một số báo cáo thông qua một

loạt các trục và mẫu và rất nhiều quy trình thủ công. Hiện họ đang tìm cách sử dụng Power BI và tự động hóa việc này. Họ không yêu cầu bất kỳ yêu cầu cụ thể nào mà muốn xem kỹ thuật có thể có trong Power BI là gì.

Mô tả dataset: Số lượng cột là đáng kinh ngạc và dữ liệu không tối ưu cho báo cáo Power BI. Có một từ điển dữ liệu được đính kèm bao gồm một số cột quan trọng nhất và ý nghĩa của chúng. Không có yêu cầu phải lập mô hình mọi thứ chỉ cần những gì bạn cảm thấy là cần thiết. Việc trích xuất tệp phẳng sang excel thường gặp khi làm việc với các hệ thống nguồn không có tích hợp trực tiếp hoặc phần mềm trung gian (kho dữ liệu). Nhiều công ty thích giữ báo cáo tách biệt với hệ thống vận hành để bảo vệ hệ điều hành khỏi bất kỳ cản trở nào về hiệu suất. Hãy đưa giải pháp và thiết kế báo cáo Power BI cho họ.

6. AdventureWorksDW2019

Lựa chọn một chủ đề trong database AdventureWorksDW2019.

Tài liệu:

https://dataedo.com/samples/html/Data_warehouse/doc/AdventureWorksDW_4/home.html

- IP: 52.250.10.176
- UID: ftu_student
- Pass: ftudb@123

7. ContosoRetailDW

Lựa chọn một chủ đề trong database ContosoRetailDW.

- IP: 52.250.10.176
- UID: ftu_student
- Pass: ftudb@123