

TRƯỜNG ĐẠI HỌC HỌC VĂN LANG  
KHOA CÔNG NGHỆ THÔNG TIN



**VAN LANG**  
UNIVERSITY



**BÁO CÁO BÀI TẬP LỚN MÔN HỌC HK241**  
**CÁC HỆ HỖ TRỢ RA QUYẾT ĐỊNH**

# **CÔNG CỤ POWER BI VÀ TRỰC QUAN HÓA**

Nhóm sinh viên thực hiện (Họ tên - Mã SV):

1. Trương Việt Vũ - 2274802011045
2. Nguyễn Phúc Nguyên - 2274802010586

GVHD: Nguyễn Thái Anh

TP. Hồ Chí Minh – năm 2024

# Mục lục

PHẦN 1: CÔNG CỤ POWER BI .....	3
1.    Giới thiệu tổng quan về Power BI. ....	3
2.    Các thành phần chính của Power BI. ....	3
PHẦN 2: TRỰC QUAN HÓA DỮ LIỆU VỚI POWER BI.....	4
1.    Giới thiệu chung về Dashboad. ....	4
1.1 Mục tiêu. ....	4
1.2 Cấu trúc.....	5
2.    Phân tích chi tiết từng thành phần của Dashboard. ....	5
2.1. Chuẩn bị các dữ liệu mới cần thiết: .....	5
2.2 Tổng số vụ tai nạn và phân chia theo giới tính.....	6
2.3 Tỷ lệ các loại sự cố xảy ra. ....	7
2.4 Tổng số tiền phải đền bù dựa theo các buổi trong ngày. ....	8
2.5 Chi phí đền bù cho các loại thiệt hại theo nhóm tuổi. ....	8
2.6 Tiền phải đền bù dựa theo hãng xe.....	9
2.7 Bản đồ số vụ tai nạn trên các thành phố. ....	10
3.    Phân tích tổng quan về xu hướng và đề xuất cải tiến.....	11
3.1 Xu hướng chung. ....	11
3.2 Đề xuất cho các biện pháp cải thiện .....	11
PHẦN 3: ỨNG DỤNG MÔ HÌNH AI ĐƠN GIẢN.....	13

# PHẦN 1: CÔNG CỤ POWER BI.

## 1. Giới thiệu tổng quan về Power BI.

Power BI là một công cụ được phát triển bởi Microsoft, được dùng để phân tích dữ liệu và hỗ trợ ra quyết định. Power BI là một trong những công cụ phổ biến và mạnh mẽ nhất hiện nay, giúp người dùng phân tích và trực quan hóa dữ liệu dễ dàng. Dưới đây là những tính năng quan trọng của Power BI:

- **Import:** Power BI trích xuất dữ liệu từ các máy chủ, trang tính Excel, tệp CSV và các cơ sở dữ liệu khác nhau. Nếu bạn nhập trực tiếp dữ liệu trong Power BI, dữ liệu đó sẽ chỉ được nén tối đa 1 GB.
- **Transform data:** Trước khi trực quan hóa dữ liệu, nên làm sạch và xử lý trước dữ liệu. Điều này có nghĩa là loại bỏ các giá trị không cần sử dụng đến hoặc bị thiếu khỏi các hàng hoặc cột.
- **Trực quan hóa dữ liệu:** Sau khi làm sạch và chuyển đổi dữ liệu, các báo cáo sẽ được tạo dựa trên các yêu cầu. Báo cáo là hình ảnh trực quan hóa dữ liệu với các bộ lọc và ràng buộc khác nhau được trình bày dưới dạng đồ thị, biểu đồ và các số liệu khác.

## 2. Các thành phần chính của Power BI.

- **Power BI Desktop** là ứng dụng được cài đặt trên máy tính cá nhân, được sử dụng để phát triển báo cáo và trực quan hóa dữ liệu. Nó cung cấp các công cụ để kết nối với nguồn dữ liệu, làm sạch, chuyển đổi dữ liệu và tạo các báo cáo tương tác bằng cách sử dụng biểu đồ, bảng số liệu. Power BI Desktop là công cụ chính để thực hiện các tác vụ phân tích trước khi xuất bản báo cáo.
- **Power Query** là công cụ tích hợp trong Power BI Desktop, cho phép người dùng thực hiện các tác vụ chuyển đổi dữ liệu, làm sạch dữ liệu và biến đổi dữ liệu trước khi tạo các báo cáo trực quan. Công cụ này hỗ trợ các thao tác như xóa dữ liệu trùng lặp, hợp nhất bảng dữ liệu, thêm cột tính toán, và nhiều thao tác khác.

- **Power BI Service (Power BI Online):** Đây là dịch vụ đám mây của Power BI, cho phép người dùng đăng tải báo cáo và dashboard từ Power BI Desktop lên và chia sẻ với những người khác. Người dùng có thể truy cập, tương tác với báo cáo, và làm việc cộng tác qua Power BI Service.

tính, độ tuổi, và hãng xe. Thông qua đó, dashboard giúp người xem hiểu rõ hơn các chi phí bảo hiểm liên quan và tình trạng tai nạn xe hơi tại các thành phố lớn.

Việc nắm bắt các xu hướng về tai nạn xe và chi phí đền bù sẽ hỗ trợ các công ty bảo hiểm và người quản lý đưa ra các quyết định chính xác hơn. Ví dụ, bằng cách xác định những nhóm đối tượng có nguy cơ cao hơn hoặc những thời điểm trong ngày tai nạn xảy ra nhiều, các biện pháp phòng ngừa và chính sách bảo hiểm có thể được tối ưu hóa, giúp giảm thiểu tổn thất và bảo vệ an toàn giao thông.

## 1.2 Cấu trúc.

Dashboard được thiết kế với bố cục trực quan bằng công cụ Power BI, các loại biểu đồ đa dạng được sắp xếp trực quan nhằm cung cấp thông tin rõ ràng và dễ tiếp cận. Các biểu đồ bao gồm biểu đồ cột, biểu đồ tròn, biểu đồ đường, và bản đồ, mỗi loại đều hiển thị các loại dữ liệu cụ thể.

## 2. Phân tích chi tiết từng thành phần của Dashboard.

### 2.1. Chuẩn bị các dữ liệu mới cần thiết:

- Cột Thời gian trong ngày (TimeOfDay):

Dùng để xác định các khoảng thời gian trong ngày như: sáng, chiều, tối. Được tạo ra từ cột incident\_hour\_of\_the\_day (giờ xảy ra sự cố trong ngày) nhằm mục đích phân chia ra các khoảng thời gian để dễ hình dung và nắm rõ hơn về xu hướng thời gian của các sự cố.

```
1 TimeOfDay =  
2 SWITCH(TRUE(),  
3   'data'[incident_hour_of_the_day] < 12, "Sáng",  
4   'data'[incident_hour_of_the_day] < 18, "Chiều",  
5   'data'[incident_hour_of_the_day] <= 23, "Tối",  
6   "Khác")
```

Hình 2: Thêm cột TimeOfDay.

- Cột Nhóm tuổi (AgeGroup):

Được tạo ra để phân chia các nhóm tuổi dựa trên độ tuổi của customer để dễ sử dụng và phân tích các hành vi dựa trên nhóm tuổi.

```

1 AgeGroup =
2 SWITCH(TRUE(),
3   'data'[age] < 25, "<25",
4   'data'[age] <= 35, "25-35",
5   'data'[age] <= 45, "36-45",
6   'data'[age] <= 55, "46-55",
7   ">55")
8

```

Hình 3: Thêm cột AgeGroup.

- Kinh độ và vĩ độ (Latitude & Longitude) :

Khi vẽ biểu đồ dạng map thì cần được cung cấp vị trí kinh độ và vĩ độ để có thể xác định chính xác vị trí của các thành phố. Được xác định dựa trên các thành phố từ cột incident\_city để đưa vào tọa độ địa lí chính xác.

<pre> 1 Latitude = IF( 2   [incident_city] = "Arlington", 38.8816, 3   IF( 4     [incident_city] = "Columbus", 39.9612, 5     IF( 6       [incident_city] = "Hillsdale", 41.7532, 7       IF( 8         [incident_city] = "Northbend", 39.3252, 9         IF( 10          [incident_city] = "Northbrook", 42.1130, 11          IF( 12            [incident_city] = "Riverwood", 41.9022, 13            IF( 14              [incident_city] = "Springfield", 39.7834, 15              BLANK() // Giá trị mặc định nếu không khớp 16            ) 17          ) 18        ) 19      ) 20    ) 21  ) 22 ) </pre>	<pre> 1 Longitude = IF( 2   [incident_city] = "Arlington", -77.1068, 3   IF( 4     [incident_city] = "Columbus", -82.9988, 5     IF( 6       [incident_city] = "Hillsdale", -85.0561, 7       IF( 8         [incident_city] = "Northbend", -84.6021, 9         IF( 10          [incident_city] = "Northbrook", -87.8125, 11          IF( 12            [incident_city] = "Riverwood", -87.6878, 13            IF( 14              [incident_city] = "Springfield", -89.6501, 15              BLANK() // Giá trị mặc định nếu không khớp 16            ) 17          ) 18        ) 19      ) 20    ) 21  ) 22 ) </pre>
---	---

Hình 4: Thêm cột Latitude và Longitude .

- Measure Số lượng các vụ tai nạn (IncidentCount):

Được xây dựng bằng cách đếm tổng số dòng của dữ liệu, được sử dụng để đếm số lượng các vụ tai nạn.

```

1 IncidentCount = COUNTROWS(data)

```

Hình 5: Thêm Measure IncidentCount

## 2.2 Tổng số vụ tai nạn và phân chia theo giới tính.

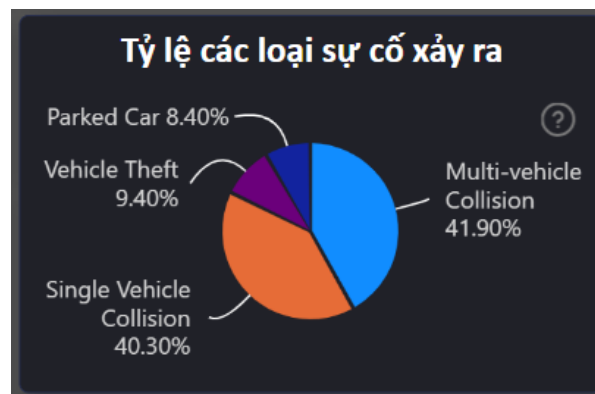
- Tổng số vụ tai nạn: Chỉ số tổng số vụ tai nạn cung cấp tổng thể về mức độ xảy ra tai nạn xe hơi. Đây là thông tin quan trọng để đánh giá mức độ xảy ra của các sự

cố giao thông và xác định xu hướng chung. Chúng tôi sử dụng thẻ “Card” giúp cho người xem dễ dàng nắm bắt được thông tin về số vụ tai nạn ngay lập tức.

- Phân tích theo giới tính: Việc phân chia các vụ tai nạn theo giới tính giúp nhận biết sự khác biệt giữa nam và nữ. Giúp các bên liên quan xem xét đưa ra các biện pháp bảo vệ hoặc các đề xuất phù hợp theo giới tính để giảm thiểu thiệt hại. Biểu đồ thanh ngang được lựa chọn giúp cho việc so sánh giữa hai giới tính (nam và nữ) vì nó trực quan và giúp người xem dễ dàng nhận biết sự khác biệt giữa các giá trị. Theo biểu đồ ta có thể thấy số lượng phụ nữ gặp phải sự cố hơn nam giới, từ đó chúng ta có thể đề xuất bán bảo hiểm xe hơi hơn vào tệp khách hàng nữ để tăng được lợi nhuận.

### 2.3 Tỷ lệ các loại sự cố xảy ra.

Các loại sự cố được chia thành Parked Car (đỗ xe), Vehicle Theft (trộm xe), Single Vehicle Collision (va chạm đơn lẻ), và Multi-vehicle Collision (va chạm nhiều phương tiện). Việc phân loại chi tiết các sự cố giúp người dùng hiểu rõ về mức độ của từng loại sự cố, từ đó có thể xác định loại tai nạn nào phổ biến nhất. Biểu đồ tròn giúp thể hiện tỷ lệ của từng loại sự cố trong tổng số các vụ tai nạn. Đây là cách tốt nhất để người xem thấy rõ phần trăm của mỗi loại sự cố.

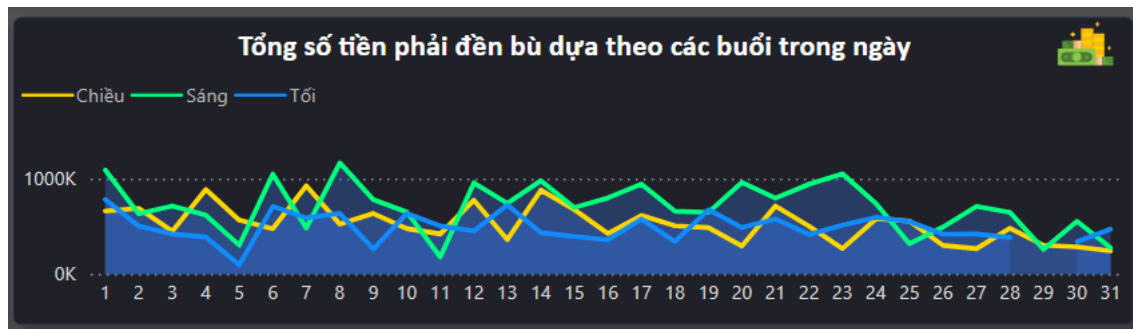


Hình 6: Biểu đồ Tỷ lệ các loại sự cố.

Dựa vào biểu đồ ta thấy xu hướng chung: Multi-vehicle Collision và Single Vehicle Collision chiếm tỷ lệ lớn nhất chiếm hơn 82%, điều này cho thấy đây là 2 loại sự cố phổ biến nhất và có thể yêu cầu các biện pháp an toàn giao thông đặc biệt để giảm thiểu.

## 2.4 Tổng số tiền phải đền bù dựa theo các buổi trong ngày.

Việc xem xét tổng chi phí đền bù dựa trên thời điểm xảy ra tai nạn (sáng, chiều, tối) giúp xác định thời gian có nguy cơ tai nạn cao. Điều này cung cấp số tiền cần phải đền bù cho các sự cố xe hơi theo khoản thời gian trong ngày. Biểu đồ đường dùng để hiển thị sự thay đổi theo thời gian trong 1 tháng, cho phép người xem nhanh chóng nhận ra các thời điểm có chi phí đền bù cao trong ngày và trong tháng.



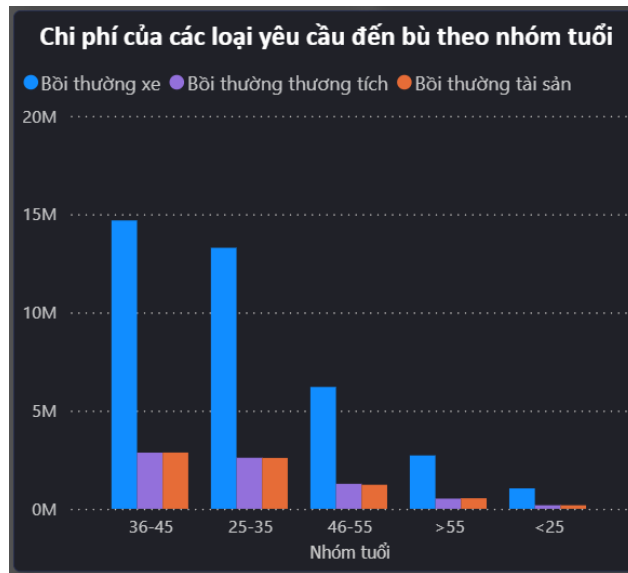
Hình 7: Biểu đồ Tổng số tiền phải đền bù dựa theo các buổi trong ngày.

Tổng quan biểu đồ ta thấy được chi phí đền bù vào buổi chiều là cao nhất lúc cao nhất có thể đạt đến hơn 1.670.000\$, điều này chỉ ra rằng thời điểm này xảy ra nhiều vụ tai nạn hoặc các tai nạn vào buổi chiều thường nghiêm trọng hơn. Còn vào buổi sáng và tối thì tuy ít hơn nhưng nhìn chung vẫn có xu hướng giảm dần khi về cuối tháng.

## 2.5 Chi phí đền bù cho các loại thiệt hại theo nhóm tuổi.

Việc phân tích chi phí của các loại yêu cầu đền bù dựa theo nhóm tuổi giúp xác định nhóm tuổi nào có nguy cơ cao và chi phí cần phải bồi thường cho các loại yêu cầu bồi thường chiếm đa số. Điều này giúp bảo hiểm định giá chính xác hơn cho từng nhóm đối tượng cũng như các loại yêu cầu bồi thường. Việc sử dụng biểu đồ cột thích hợp để thể hiện sự khác biệt về chi phí của các loại yêu cầu cần đền bù giữa các nhóm tuổi, giúp người xem dễ dàng so sánh sự khác biệt.





Hình 8: Biểu đồ Chi phí của các loại yêu cầu đền bù theo nhóm tuổi.

Nhóm tuổi 36-45 và 25-35 có chi phí yêu cầu bồi thường cao nhất đạt gần 15.000.000\$ mỗi nhóm, và đa số đều có yêu cầu chi phí cao cho bồi thường xe ở mọi nhóm tuổi, điều này có thể chỉ ra rằng các đối tượng trong độ tuổi này thường tham gia giao thông nhiều và dễ xảy ra tai nạn hơn, bên cạnh đó yêu cầu bồi thường xe luôn là vấn đề chiếm nhiều chi phí bồi thường nhất ở đây. Độ tuổi bé hơn 25 có thể vẫn chưa có đủ kinh tế để mua xe nên chi phí chiếm khá thấp. Nhìn chung chi phí yêu cầu bồi thường có xu hướng giảm dần theo độ tuổi người sử dụng từ 36 tuổi đến hơn 55 tuổi.

## 2.6 Tiền phải đền bù dựa theo hãng xe.

Chi phí đền bù theo hãng xe giúp các công ty bảo hiểm xác định được mức độ rủi ro của từng dòng xe. Các xe thuộc hãng có chi phí đền bù cao có thể có mức phí bảo hiểm cao hơn. Biểu đồ dạng phễu (Funnel) giúp dễ dàng so sánh chi phí giữa các hãng xe, cung cấp



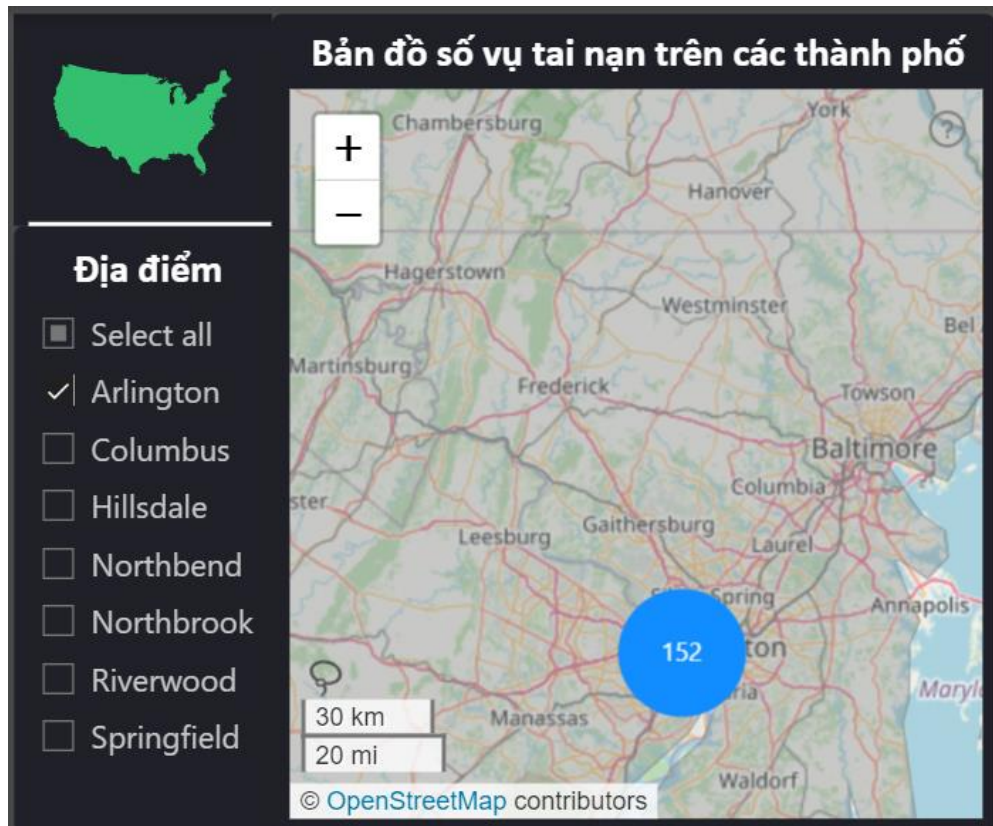
Hình 9: Biểu đồ tiền bồi thường dựa trên hãng xe.

một cái nhìn nhanh về hãng nào có mức chi phí đền bù cao nhất, cũng như sự phân bố tiền bồi thường cho các hãng xe nổi tiếng.

Ta thấy hãng xe Dodge chi phí đền bù cao, gần 4.500.000\$. Điều này có thể do xe thuộc hãng này dễ bị tai nạn hoặc chi phí sửa chữa cao hơn. Từ đó ta có thể thương lượng lại giá bảo hiểm với hãng Dodge cũng như tập trung vào tệp khách hàng sử dụng xe từ hãng này và các hãng khác cũng có chi phí cao trong biểu đồ.

## 2.7 Bản đồ số vụ tai nạn trên các thành phố.

Hiển thị số vụ tai nạn trên bản đồ giúp xác định các thành phố có tỷ lệ gặp cao, từ đó chúng ta có thể tập trung cải thiện quy trình bảo hiểm và bồi thường vì sẽ có số lượng khách hàng gặp sự cố nhiều hơn những nơi khác giúp nâng cao trải nghiệm của người dùng. Bản đồ là công cụ tốt nhất để thể hiện thông tin địa lý, cho phép người xem nhận diện nhanh chóng các khu vực có nhiều tai nạn.



Hình 10: Ví dụ Bản đồ số vụ tai nạn trên thành phố Arlington

Các thành phố có số vụ tai nạn cao là những khu vực cần tăng cường các biện pháp phục vụ khách hàng. Cần được chú trọng hơn những khu vực khác với ít sự cố hơn. Thực hiện quảng bá và hướng dẫn những khách hàng tiềm năng trong khu vực thường xuyên xảy ra sự cố xe hơi này để nâng cao độ hiệu quả.

### 3. Phân tích tổng quan về xu hướng và đề xuất cải tiến.

#### 3.1 Xu hướng chung.

Tổng quan các xu hướng chính cho thấy phụ nữ có tỉ lệ gặp sự cố cao hơn đàn ông, tai nạn xảy ra nhiều hơn vào buổi chiều, chi phí đền bù cao hơn ở nhóm tuổi từ 25-45. Multi-vehicle Collision là loại tai nạn phổ biến nhất, các hãng xe phổ thông như Dodge thường gặp nhiều sự cố hơn. Đa số các loại yêu cầu đền bù đều là đền bù về xe.

#### 3.2 Đề xuất cho các biện pháp cải thiện

1. Tăng cường chính sách bảo hiểm và biện pháp phòng ngừa:

- Điều chỉnh chính sách cho từng nhóm khách hàng cụ thể: Dựa trên dữ liệu, phụ nữ và nhóm tuổi từ 25-45 có tỷ lệ gặp sự cố cao hơn so với các nhóm khác. Để giảm thiểu rủi ro này, công ty bảo hiểm có thể thiết kế các gói bảo hiểm đặc thù hoặc đưa ra mức ưu đãi phù hợp cho các nhóm này, nhằm khuyến khích họ tham gia các chương trình đào tạo hoặc kiểm tra sức khỏe định kỳ để nâng cao ý thức lái xe an toàn.
- Triển khai các chương trình đào tạo kỹ năng lái xe: Đặc biệt trong giờ cao điểm, như buổi chiều, là thời điểm có tỷ lệ tai nạn cao. Công ty có thể hợp tác với các cơ quan chức năng để tổ chức các khóa học, hội thảo nhằm nâng cao kỹ năng và ý thức lái xe an toàn, đồng thời có thể đưa ra các ưu đãi phí bảo hiểm cho khách hàng tham gia chương trình đào tạo.

## *2. Điều chỉnh mức phí bảo hiểm theo hãng xe:*

- Tăng mức phí bảo hiểm cho các dòng xe có nguy cơ cao: Các xe của hãng Dodge được xác định có chi phí đền bù cao, điều này có thể phản ánh mức độ rủi ro cao hơn so với các hãng khác. Công ty bảo hiểm có thể áp dụng mức phí cao hơn hoặc các điều khoản bảo hiểm bổ sung cho dòng xe này để bù đắp chi phí đền bù cao.
- Hợp tác với các nhà sản xuất xe để cải tiến độ an toàn: Công ty có thể xây dựng các chương trình hợp tác với nhà sản xuất xe như Dodge để triển khai các biện pháp tăng cường an toàn cho xe của hãng, bao gồm việc cải tiến tính năng an toàn và cung cấp các khóa hướng dẫn lái xe an toàn cho chủ xe. Điều này không chỉ giúp giảm rủi ro tai nạn mà còn có thể tăng sự tin tưởng của khách hàng đối với hãng bảo hiểm.

## *3. Tăng cường hỗ trợ tại các thành phố có tỷ lệ tai nạn cao:*

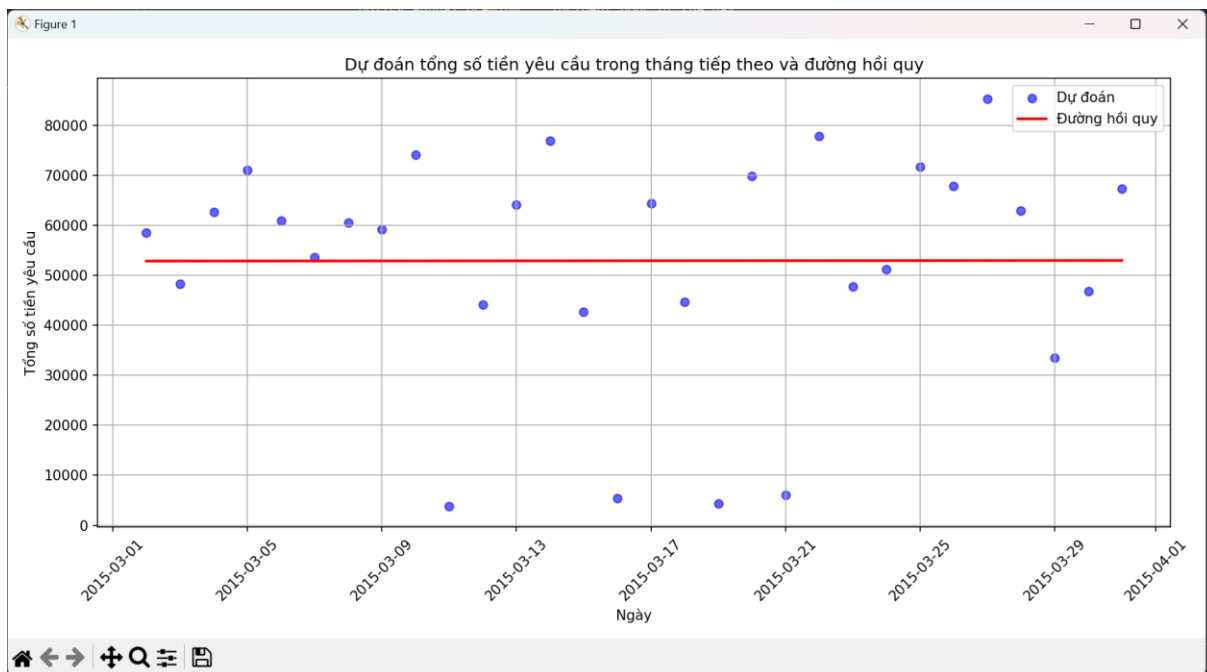
Cải thiện dịch vụ bồi thường và hỗ trợ khách hàng tại các khu vực rủi ro cao: Với những thành phố có tỷ lệ tai nạn cao như được thể hiện qua bản đồ trong dashboard, công ty có thể đầu tư thêm vào dịch vụ chăm sóc khách hàng và tối ưu hóa quy trình bồi thường để hỗ trợ khách hàng nhanh chóng và hiệu quả hơn.

### PHẦN 3: ỨNG DỤNG MÔ HÌNH AI ĐƠN GIẢN.

Nhận thức được sự quan trọng cũng như tầm ảnh hưởng của mức phí dùng để bồi thường thiệt hại. Nhóm đã tiến hành training AI để có thể đưa ra được dự báo về số tiền dự đoán cần dùng để đền bù thiệt hại cho khách hàng trong tháng tiếp theo (tức tháng 3 năm 2015) theo từng ngày để có thể sớm đưa ra những chính sách tốt nhất để nâng cao chất lượng phục vụ khách hàng.

Nhóm đã sử dụng mô hình Linear Regression (Hồi quy tuyến tính) để tiến hành dự đoán cho số tiền đền bù theo từng ngày trong tháng tiếp theo dựa trên tập dữ liệu ban đầu đã được thông qua các bước làm sạch dữ liệu bằng thư viện Pandas của Python.

Sau khi đã thu được kết quả, nhóm đã sử dụng thêm thư viện Matplotlib để trực quan hóa dữ liệu theo mô hình Linear Regression để có một cái nhìn tổng quan dự đoán:



Nhìn chung số tiền bồi thường có xu hướng luôn nằm gần đường hồi quy với giá trị từ khoảng 40.000\$ đến 80.000\$ trong tháng tiếp theo. Bên cạnh đó cũng có những ngày có thể rất thấp khi dưới 10.000\$. Dựa vào biểu đồ dự đoán mà chúng ta có thể đưa ra những quyết định nhằm mục đích nâng cao chất lượng công việc cũng như tránh đi những trường hợp ngoài ý muốn.